NBER WORKING PAPER SERIES

GERMAN LONG-TERM HEALTH INSURANCE: THEORY MEETS EVIDENCE

Juan Pablo Atal
Hanming Fang
Martin Karlsson
Nicolas R. Ziebarth

Working Paper 26870
http://www.nber.org/papers/w26870

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2020, Revised May 2023

Previously circulated as “Long-Term Health Insurance: Theory Meets Evidence.” We are grateful to the editor, Harald Uhlig, and two anonymous referees for detailed comments that significantly improved the paper. Siqi Li provided outstanding research assistance. We also thank Konstantin Beck, Paola Bertoli, Rudi Blankert, Friedrich Breyer, Florian Buchner, Keith Ericson, Colleen Carey, Mariacristina De Nardi, David Dillenberger, Stefan Felder, Sebastian Fleitas, Eric French, Leora Friedberg, Paul Grieco, Ben Handel, Glenn Harrison, Jim Heckmann, Kate Ho, Mathias Kifmann, Amanda Kowalski, Dirk Krueger, Ethan Lieber, Tony Lo Sasso, Claudio Lucarelli, Maria Polyakova, John Rust, Andreas Ryll, Holger Strulik, Nicholas Tilipman, JürgenWasem, Michael Whinston, JoachimWinter, Moto Yogo, Jin Zhou, and Peter Zweifel as well as seminar participants at Boston University, Columbia University, Indiana University, Georgia State University, Georgetown University, Johns Hopkins University, Lund University, Princeton University, University of Chicago, University of Pennsylvania, University of Maryland, University of Wisconsin-Madison, and conference participants at the American Society of Health Economists (2018), Annual Meeting of the Southern Economic Association (2018), Essen Health Conference (2019), Congress of the International Health Economics Association (iHEA 2019), GSE Forum in Microeconometrics (2019), Health Economics Working Group Meeting of the German Economic Association (Ausschuss für Gesundheitsökonomie, 2018), NBER Insurance Working Group Conference (2019), New York Area Health Economics Day (2019) and the North American Winter Meeting of the Econometric Society (2022) for very helpful comments and suggestions. We thank representatives of the German Association of Private Health Insurers and Frank Bowert for invaluable comments on how well our modeling of German LTHI reflects actuarial practices, and we thank Friedrich Breyer for connecting us with representatives from the PKV which provided us the claims data.We have no financial interests that would constitute any conflict of interests with this research. Generous funding by the German Federal Ministry of Education and Research (FKZ: 01EH1602A) is gratefully acknowledged. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

German Long-Term Health Insurance: Theory Meets Evidence
Juan Pablo Atal, Hanming Fang, Martin Karlsson, and Nicolas R. Ziebarth
NBER Working Paper No. 26870
March 2020, Revised May 2023
JEL No. G22,I11,I18

ABSTRACT

By insuring policyholders against contemporaneous health expenditure shocks and future reclassification risk, long-term health insurance contracts are a viable alternative to community-rated short-term contracts with an individual mandate. German long-term health insurance (GLTHI) is the largest market for private long-term health insurance contracts in the world. It features a simple design with initial risk-rating followed by guaranteed-renewable constant premiums over the lifecycle. We estimate the key ingredients of a life-cycle model to assess the welfare effects of the GLTHI contract and compare them to the optimal contract. This comparison provides further lessons about the trade-offs of long-term health insurance design.

Juan Pablo Atal
University of Pennsylvania
Department of Economics
The Ronald O. Perelman Center for
Political Science and Economics
133 South 36th Street
Suite 150
Philadelphia, PA 19104
ataljp@econ.upenn.edu

Hanming Fang
Department of Economics
Ronald O. Perelman Center
for Political Science and Economics
University of Pennsylvania
133 South 36th Street, Suite 150
Philadelphia, PA 19104
and NBER
hanming.fang@econ.upenn.edu

Martin Karlsson
University of Duisburg-Essen
mkmkarlsson@gmail.com

Nicolas R. Ziebarth
Cornell University
Cornell Jeb E. Brooks School of Public Policy
Department of Economics
2218 MVR
Ithaca, NY 14853
and ZEW Mannheim
and also NBER
nrz2@cornell.edu

1 Introduction

Short-term health insurance contracts can expose policyholders to potentially large premium fluctuations ("reclassification risk") which can lead to significant welfare losses (Diamond, 1992; Cochrane, 1995). Consequently, for decades, academics and policymakers have debated how to regulate short-term health insurance markets. The standard policies, such as community-rated premiums and guaranteed issuance regulations, strive to avoid reclassification risk, uninsurance, and unaffordable premiums for sick individuals (Claxton et al., 2017; Cole et al., 2019). However, these policies also imply a trade-off with unintended consequences such as adverse selection. These are then typically addressed through individual mandates, premium subsidies, or both (cf. Akerlof, 1970). For example, the Affordable Care Act (ACA), enacted in the United States in 2010, established health insurance exchanges for non-group plans featuring community rating, an individual mandate, and premium subsidies as its three main pillars (Aizawa and Fang, 2020). However, the controversies of mandates and the costs associated to subsidies continue to call for examining alternative solutions.[1]

Long-term (also known as *dynamic*) health insurance contracts offer a fundamental alternative to short-term contracts, and can provide reclassification risk insurance without triggering adverse selection. They do so without mandates or subsidies. Instead, long-term contracts leverage individuals' intertemporal incentives to insure future premium risk with a schedule of "frontloaded" premiums (Pauly et al., 1995, 1999; Patel and Pauly, 2002; Pauly and Lieberthal, 2008). Front-loaded premiums based on initial risk-rating generate "lock-in", which help insure reclassification risk while also preventing adverse selection even if individuals are free to lapse policies ("one-sided commitment"). In particular, Ghili et al. (2022) derive the long-term health insurance contract that optimally balances reclassification risk with the detrimental effects on intertemporal consumption smoothing generated by frontloading, in a context with one-sided commitment and borrowing constraints. Despite their theoretical appeal, little is known about real-world applications of long-term contracting in health insurance, and how their implementation relates to the theoretical prescriptions.

In this paper, we provide the following main contributions to the literature. First, we bridge the theory of dynamic health insurance contracts with their most important real-world application: German Long-Term Health Insurance (GLTHI). We show that, in spite of their simple design, the existing GLTHI contracts share several features with the optimal dynamic contract characterized by Ghili et al. (2022). Moreover, the GLTHI contract design coincides with that of the optimal dynamic contract for individuals with a flat lifecycle income profile. Second, we evaluate welfare under this

[1]The ACA originally introduced an individual "mandate" through imposing a tax penalty for people without insurance (with some exceptions). This was the subject of many controversies and legal battles. The tax penalty was later set to zero in 2019.

single, "flat-income-optimal" long-term contract. We then compare it to welfare under the optimal contracts tailored to the increasing income profiles observed in the data. In doing so, we highlight under which circumstances the performance of optimal contracts is robust to the simplifications of the German design.

Worldwide, besides Chile, Germany is one of the two countries with comprehensive long-term health insurance contracts. Several features make the German market especially appealing to study. First, it is the largest and oldest individual private long-term health insurance market in existence; its roots go back to 1883. Currently, it covers almost 9 million individuals. Second, contracts in this market are purely financial—with no differentiation in the provider network across insurers or plans. The absence of networks makes it well suited for a comparison with the theory of optimal dynamic pricing in health insurance (which abstracts away from such differentiation across plans). Third, after an initial risk rating, contracts follow the simple principle of guaranteed renewability with constant and income-independent premiums as long as the policyholder decides to renew. This simple design is particularly appealing from a policy perspective.

We begin by presenting the main principles of GLTHI, where we focus on establishing their commonalities with the principles of optimal contracts. Importantly, we show that when income is flat over the lifecycle, the GLTHI contracts coincide with the optimal dynamic contracts derived by Ghili et al. (2022). Using data from one of the largest German private insurer, we find that the theoretically optimal contract for flat lifecycle incomes matches the empirically observed GLTHI premiums very well. More generally, for non-flat lifecycle incomes, we show that the simple GLTHI design provides more reclassification risk insurance and charges higher premiums early in life (*frontloading*) than the optimal contract would. This results in less consumption smoothing over the lifecycle than under the optimal dynamic contract.

Next, we estimate a lifecycle model to assess welfare under the German design of long-term health insurance contracts. Our main goal is to quantify the welfare gap when using the simple GLTHI design instead of offering menus of contracts that optimally trade off consumption smoothing with reclassification risk for each income profile. To do so, we rely on a unique panel of claims data from one of the largest German private health insurers. The data cover 620 thousand policyholders over 7 years, spanning all age groups and all of the 16 German federal states.[2] Key to our analysis, we propose and implement a novel health risk classification method to parsimoniously model individuals' expected health risks and their evolution over the lifecycle, while recovering relevant features of typical health expenditure distributions. We then combine the claims data with income data from

[2]The oldest policyholder is 99 years old and one policyholder has been a client for 86 years, illustrating the long-term and lifecycle dimension of these policies.

Germany's Socieoconomic Panel Longitudinal Survey (SOEP).

Within a wide range of assumptions, we find that welfare under the GLTHI design—as measured by the certainty-equivalent yearly consumption—is close to 96 percent of what the optimal contract achieves. The welfare loss of using a constant lifecycle premium profile is partly compensated by less reclassification risk than in the income-dependent optimal contract. We find welfare differences of a similar magnitude if we consider (i) a wide range of risk aversion, (ii) different functional forms for utility, including Epstein-Zin recursive preferences, (iii) private savings, (iv) an initial health status that matches the German population as a whole (instead of the privately insured), and (v) U.S. income profiles.

We finish by discussing the circumstances under which the performance of optimal contracts is robust to the simplifications of the German design. In theory, the skewness or health care expenditures may result in large welfare losses from reclassification risk under a series of actuarially-fair short-term contracts. In such cases, optimal long-term contracts are particularly effective in improving upon short-term contracts: The welfare gains from reducing catastrophic reclassification risk are large relative to the losses introduced by premium frontloading. Our findings imply that, in those circumstances, the particular balance between reclassification risk and frontloading of the flat-income optimal contract achieves much of the benefits of optimal long-term contracting. However, if catastrophic losses under short-term contract were more limited *via* other means—for instance, due to government social safety-net programs against catastrophic losses—the gains from optimal long-term contracting relative to short-contracting would be smaller. The GLTHI contract would then recoup a smaller fraction of those incremental gains. Overall, our results suggest that the good performance of optimal long-term contracts in dealing with large reclassification risk is robust to the use of flat income profiles. We see this as a virtue of the theory of dynamic contracting in health insurance.

Our results have policy implications for the U.S. and elsewhere. The case of Germany shows that long-term contracts are a viable alternative to regulated short-term contracts for insuring reclassification risks. It also shows that the simple "flat-income" version of optimal long-term contracts implemented in Germany—a single long-term contract with constant premium guarantees—can recoup much of the overall potential gains of optimal long-term contracting, especially when reclassification risk is not already insured by other means.

This paper contributes to several strands of literature. First, it contributes to the literature on dynamic contracts for which vast theoretical work but relatively little empirical evidence exists. Pauly et al. (1995) propose a "guaranteed-renewable" contract with a pre-specified path of premiums that fully eliminates adverse selection and reclassification risk. Similarly, Cochrane (1995) proposes a scheme of severance payments, made after the realization of health shocks, which provides full insurance against reclassification risks. Harris and Holmstrom (1982) and Krueger and Uhlig (2006) study the properties of competitive long-term contracts that insure agents against income risk under one-sided commitment when agents receive endogenous outside offers. Hendel and Lizzeri (2003) and Ghili et al. (2022) show that the optimal dynamic insurance contract only partially insures reclassification risk, because fully eliminating reclassification risks requires large front-loaded payments, which prevents consumption smoothing over the lifecycle. Importantly, we build heavily on Ghili et al. (2022), as we use their characterization of the optimal long-term health insurance contract, and perform similar welfare comparisons. One key addition is to show how the most important real-world application of long-term health insurance contracts corresponds to a version of their theoretically optimal contract—the one that would emerge under flat income profiles—and to quantify the welfare gap from such a sub-optimal implementation of their theory. We also characterize and quantify the one-period state-contingent Arrow Securities that can be used to implement the GLTHI as well as the optimal contract (Krueger and Uhlig, 2006). Arrow Securities have the advantage of fostering insurer competition by overcoming the detrimental competitive effects of lock-in in long-term contracts.[3]

Second, several papers, including Hendel and Lizzeri (2003), Herring and Pauly (2006), Finkelstein et al. (2005), Fleitas et al. (2018), and Atal (2019), investigate empirically the workings of long-term contracts in different contexts. Our paper contributes to this literature by contrasting the empirical applications of health insurance contracts with their optimal design over the lifecycle. We also introduce a novel data-driven method of discrete classification of health risks. We base our method on an explicit objective function proposed in the actuarial science literature that maximizes the differentiation between risk classes, invoking the properties of *homogeneity* and *separation* in risk classification (see Finger, 2006). By better capturing the skewness of healthcare expenditures, our method improves upon the mostly *ad hoc* methods for discretizing health risks used in the literature. Importantly, accurately modeling the right tail of health expenditure distributions allows us to highlight the workings of long-term contracts in insuring individuals against large catastrophic risks.

Finally, our paper relates to previous work on the German long-term health insurance market. Hofmann and Browne (2013) describe the actuarial principles of premium calculation, and document that GLTHI contracts involve premium frontloading. They also show that lapsation decisions are

[3] We abstract from the effects of long-term health insurance on health transitions in this paper. See Cole et al. (2019) for a quantitative dynamic model of health investments and insurance that studies the short and long-term effects of community-rated health insurance.

consistent with its incentive structure. We contribute by explicitly comparing the features of GLTHI contracts with the optimal long-term insurance contract design, by showing that it coincides with the optimal contract under flat income profiles, and by evaluating its welfare consequences. Christiansen et al. (2016) empirically study determinants of lapsing and switching behavior. Baumann et al. (2008) and Eekhoff et al. (2006) discuss the potential effects of higher switching rates on market competition if the capital accumulated through front-loaded payments ("old-age provisions") were to be made portable across insurers. While these two papers discuss a hypothetical reform, Atal et al. (2019) theoretically and empirically study the effects of the actual 2009 portability reform on switching behavior.

2 Institutional Background

Germany has a two-tier health insurance system. Since 2009, Germany has an individual mandate and every resident must have health insurance (§193 *Versicherungsvertragsgesetz*).[4] The first tier consists of public health insurance or *Gesetzliche Krankenversicherung (GKV)* and covers 90 percent of the population of 84 million. Every employee whose income is below the politically defined federal threshold (*Versicherungspflichtgrenze*) of €66,600 in 2023 (about $70K), is *mandatorily* insured with the public multi-payer system. GKV policyholders can choose between 97 private non-profit sickness funds or *Gesetzliche Krankenkassen* (Bünnings et al., 2019; GKV-Spitzenverband, 2022). Premiums consist of income-dependent contribution rates (employees and employers each pay about eight percent of the gross wage, up to a cap) for a standardized benefit package with very little cost-sharing. Germany was the first country in the world to introduce such a public health insurance system in 1883 (*Gesetz betreffend der Krankenversicherung der Arbeiter*), the basic structure of which has remained intact since its inception.

For historical reasons, select population subgroups can *permanently* opt out of the GKV system and join the second tier: private health insurance or *Private Krankenversicherung (PKV)*, which covers the remaining ten percent or 8.7 million policyholders (Association of German Private Healthcare Insurers, 2022). The PKV covers three main population subgroups: the (a) self-employed; (b) high-income earners with gross incomes above the *Versicherungspflichtgrenze*, and (c) civil servants. They can decide to leave the public GKV system permanently, join the PKV and sign a private individual long-term health insurance contract for the rest of their lives (Hullegie and Klein, 2010; Polyakova,

[4]Despite the two tiers, the uninsurance rate is low at around 0.1 percent (Statistisches Bundesamt, 2020). Further institutional details about the two-tier German health insurance system are in Schmitz and Ziebarth (2017) (English) or Henke (2007) (German).

2016; Panthöfer, 2016).[5] To prevent individuals from strategically switching back and forth, the decision to enter the private market is essentially a lifetime decision. The basic principle is: "Once privately insured, always privately insured" (Schencking, 1999; Innungskrankenkasse Berlin Brandenburg, 2018). Appendix A1 discusses the institutional specifics and the empirical evidence on this principle.

Besides Chile (cf. Atal, 2019), Germany is the only country in the world with an existing private long-term health insurance market. The German Long-Term Health Insurance (GLTHI) market consists of 48 private insurers that sell *comprehensive* as well as *supplemental* coverage (Association of German Private Healthcare Insurers, 2020). The focus of this paper is comprehensive coverage.[6]

Provider Networks. Provider networks and "Managed Care" are unknown in both the public GKV and private PKV systems in Germany; that is, in either system, enrollees have the free choice of providers. Moreover, in both systems, reimbursement rates are centrally set and do not vary by insurers or health plans. While reimbursement rates for inpatient care are identical in both systems, they are about twice as high for outpatient care in the PKV. These structurally higher outpatient reimbursement rates result in significantly shorter waiting times for the privately insured (cf. Werbeck et al., 2021).[7] Because they do not negotiate rates or build provider networks, private insurers mainly customize health plans and process, scrutinize, and deny claims. Thus, the GLTHI contract primarily constitutes a pure financial contract similar to insurance markets such as life insurance (Fang and Kung, 2020). This specific feature substantially simplifies the welfare analysis of GLTHI contracts and its comparison with the optimal long-term contracts.[8]

One-Sided Commitment and Guaranteed Renewability. When individuals apply for a GLTHI contract, insurers have the right to deny applicants with bad risks or impose pre-existing condition clauses. However, once contracts are signed, insurers cannot terminate them. GLTHI contracts are not annual contracts, but *permanent lifetime contracts* without an end date. In other words, GLTHI contracts are guaranteed renewable over the lifecycle. However, policyholders can terminate these permanent contracts and switch insurers. Thus, the GLTHI is a market with one-sided commitment.

[5]Civil servants represent a special case as they receive about 50% of their insurance coverage from the states or federal government as their employer—and typically purchase PKV contracts for the remaining part.

[6]While comprehensive plans are not, supplemental plans are also available as group policies, primarily through private employers (*Betriebliche Krankenzusatzversicherung*). However, the market is relatively small and only counted 883,400 covered employees for the whole of Germany in 2019 (Association of German Private Healthcare Insurers, 2020).

[7]Other unintended consequences could be overtreatment of the privately insured and explicit discrimination such as separate practice waiting rooms or telephone numbers for GLTHI policyholders.

[8]Compared to public insurance, one could argue that private markets and contracts are less prone to government regulatory risk. Koijen et al. (2016) study the impact of such "government-induced profit risk" on the demand of investors for what they refer to as "medical innovation premium."

It is very common that policyholders keep their GLHTI contract until they die: Medicare does not exist in Germany.

Premium Calculation. Whereas the initial GLTHI premium is risk-rated, all subsequent premium increases are community-rated at the plan level. In fact, premiums are calculated under the basic principle of a constant premium guarantee as long as the individuals remains in the contract, sufficient to cover expenses while the individual remains in the contract. As this is a central aspect of our analysis, Section 3 discusses this issue. Appendix A2 provides additional details.

3 Lifecycle Premiums in GLTHI

3.1 Formal Derivation of GLTHI Lifecycle Premiums

This section provides a formal treatment of the main principles guiding GLTHI lifecycle premiums: risk-rated initial premiums followed by guaranteed renewability at the same premium as a long as the individual stays in the contract. Appendix A2 provides a graphical representation and additional details.

Let $P_t(\xi_t)$ be the initial premium when first signing a GLTHI contract in period t. As GLTHI contracts are individually underwritten at inception, $P_t(\xi_t)$ depends on individuals' health status in year t, ξ_t. We assume that $\xi_t \in \Xi$, where Ξ is a finite set of health states. In subsequent periods, each contract is guaranteed-renewable at the same premium. As such, individuals who sign a contract in period t can renew the contract for the same premium, $P_t(\xi_t)$, in all periods between $t+1$ and T, regardless of the evolution of their health status. However, they can lapse the existing contract and sign a new one if a new contract is available at a lower premium.

The contract breaks even in equilibrium, given premium $P_t(\xi_t)$ and the subsequent lapsation decisions of the enrollee. Thus, the equilibrium levels of $P_t(\xi_t)$ can be expressed as the solution to a fixed-point problem in which $P_t(\xi_t)$ covers exactly the expected claims of policyholders who *stay* in the contract at premium $P_t(\xi_t)$.

We solve for $P_t(\xi_t)$ recursively, starting from the last period, $t = T$. In the last period T, there is no uncertainty regarding future health status and future lapsation. Let m_t denote health care expenditures in period t. Assuming full coverage, it follows that $P_T(\xi_T) = \mathbb{E}(m_T|\xi_T)$.

To calculate the equilibrium premium in $t < T$, we need to consider *endogenous lapsation*. An interesting and practically convenient feature of the GLTHI contract is that policyholders will lapse their current contract if and only if—given the evolution of their health—they can obtain a lower premium than their current guaranteed-renewable premium $P_t(\xi_t)$. Formally, lapsing a contract signed in $t < T$ at the risk-rated premium $P_t(\xi_t)$ occurs at the first $\tau > t$ such that $P_\tau(\xi_\tau) < P_t(\xi_t)$.[9]

It may be surprising that policyholders' lapsation decisions do *not* depend on the curvature of their utility function. To understand this result, it is important to note that the difference in policyholders' continuation value from holding two long-term contracts only depends on the premium difference. This is because the other determinants of the continuation value—namely, health transitions and income dynamics—are independent of the long-term contract. Moreover, while the *level* of the value difference depends on the curvature of the utility function, the *sign* of the difference does not.[10]

Remark 1 *The lapsation decision under GLTHI is only driven by a comparison between the current guaranteed renewable premium $P_t(\xi_t)$, and the premium of a new, alternative, contract $P_\tau(\xi_\tau)$. Neither risk aversion nor income plays a role in the lapsation decision under GLTHI. As the GLTHI is a pure financial contract, differences in provider networks do not drive the lapsation decision.*

For a given $t < T$ and $\tau > t$, we denote $\mathbf{P}_{t+1}^{\tau} \equiv \{P_{t+1}(.), ..., P_\tau(.)\}$ as the set of guaranteed premiums from $t+1$ to $t+\tau$. We can then recursively write the break-even GLTHI lifecycle premium for new policyholders in period t with health state $\xi_t \in \mathcal{Z}$, denoted by $P_t(\xi_t)$, as follows:

$$P_t(\xi_t) = \frac{\mathbb{E}(m_t|\xi_t) + \sum_{\tau>t}^{T} \sum_{z\in\mathcal{Z}} \delta^{\tau-t}\mathbb{E}(m_\tau|z) \times q_\tau(z|\xi_t, \mathbf{P}_{t+1}^{\tau}, P_t(\xi_t))}{1 + \sum_{\tau>t}^{T} \sum_{z\in\mathcal{Z}} \delta^{\tau-t} \times q_\tau(z|\xi_t, \mathbf{P}_{t+1}^{\tau}, P_t(\xi_t))}, \tag{1}$$

where the first element of the numerator, $\mathbb{E}(m_t|\xi_t)$, is expected health care expenditures in period t, given ξ_t; the second element of the numerator is the sum of the expected future health care expenditures over all remaining life years from t to T. Expected future health care expenditures are discounted at rate δ, with future spending at period τ weighted by $q_\tau(z|\xi_t, \mathbf{P}_{t+1}^{\tau}, P_t(\xi_t))$, the probability that (i) $\xi_\tau = z$ and (ii) the policyholder does not lapse (or die) between periods t and τ, given the subsequent equilibrium premiums $\mathbf{P}_{t+1}^{\tau}$. These expected lifecycle expenditures are then divided by the expected number of years in contract.[11] In other words, the GLTHI lifecycle premium offered in period t to an individual in state ξ_t, $P_t(\xi_t)$, equals the annualized expected health care expenditures

[9]Note that we abstain from switching costs, and from horizontal differentiation across plans. Horizontal differentiation across plans tends to be minor because, as we explained in Section 2, GLTHI is a pure financial contract.

[10]This argument also applies when policyholders' preferences are not time separable, e.g., if they have Epstein-Zin preferences (Epstein and Zin, 1989).

[11]Of course, $q_\tau(z|\xi_t, \mathbf{P}_{t+1}^{\tau}, P_t(\xi_t))$ depends on the evolution of the health status $\xi_{t+1}, ..., \xi_\tau$ and death, conditional on current health status ξ_t. We describe how we model the evolution of health status and its implications for health expenditures in Section 5.

starting from t and while the individual remains in the contract, considering the individual's health status at the inception of the contract.

Note that these lifecycle premiums do not maximize any *ex ante* consumer objective function; conceptually, they are *not* designed to maximize any welfare criterion.

Remark 2 *The GLTHI equilibrium premiums are recursively determined by Equation* (1)*. They do not depend on the policyholder's utility function or lifecycle income profile. Therefore, the GLTHI premiums do not depend on education or other determinants of lifecycle income profiles.*

3.2 Comparison to the Optimal Dynamic Contract

Under the assumption of time-separable preferences and risk aversion, and one-sided commitment by insurers, the optimal dynamic health insurance contract as derived by Ghili et al. (2022) consists of consumption guarantees, $\bar{c}_t(\xi_t, \mathbf{y}_t^T)$. These depend not only on individuals' health but also on a vector of their current and future incomes $\mathbf{y}_t^T \equiv \{y_t, y_{t+1}, ...y_T\}$.

Analogous to the GLTHI lifecycle premium calculation, $\bar{c}_t(\xi_t, \mathbf{y}_t^T)$ can be solved by backwards induction. Specifically, the consumption guarantee in period T is given by $\bar{c}_T(\xi_T, y_T) = y_T - \mathbb{E}(m_T|\xi_T)$. For any $t < T$ and $\tau > t$, denote the set of future equilibrium consumption guarantees $\bar{\mathbf{c}}_{t+1}^{\tau} \equiv \{\bar{c}_{t+1}(.), ..., \bar{c}_{\tau}(.)\}$. An algebraic reformulation of Lemma (2) in Ghili et al. (2022) implies that the equilibrium break-even consumption guarantee under the optimal dynamic contract for an individual purchasing a long-term optimal contract at time t under health status ξ_t is recursively determined by:

$$\bar{c}_t(\xi_t, \mathbf{y}_t^T) = \frac{y_t - \mathbb{E}(m_t|\xi_t) + \sum\limits_{\tau>t}^{T} \sum\limits_{z\in\Xi} \delta^{\tau-t}(y_\tau - \mathbb{E}(m_\tau|z)) \times q_\tau(z|\xi_t, \bar{\mathbf{c}}_{t+1}^{\tau}, \bar{c}_t(\xi_t, \mathbf{y}_t^T))}{1 + \sum\limits_{\tau>t}^{T} \sum\limits_{z\in\Xi} \delta^{\tau-t} \times q_\tau(z|\xi_t, \bar{\mathbf{c}}_{t+1}^{\tau}, \bar{c}_t(\xi_t, \mathbf{y}_t^T))}, \tag{2}$$

where $q_\tau(z|\xi_t, \bar{\mathbf{c}}_{t+1}^{\tau}, \bar{c}_t(\xi_t, \mathbf{y}_t^T))$ is, with some slight abuse of notation, the probability that (i) $\xi_\tau = z$, and (ii) the individual does not lapse (or die) between periods t and τ, given the set of future equilibrium consumption guarantees $\bar{\mathbf{c}}_{t+1}^{\tau}$.[12]

The optimal and the GLTHI contract have a several commonalities. Both contracts break even in expectation. Both contracts entail frontloading to limit (but not eliminate) reclassification risk. Both contracts are risk-rated at inception and provide some form of renewable guarantee. While the GLTHI establishes a constant *premium guarantee* that *decreases* when individuals could get a better

[12]This characterization of the optimal dynamic contract is independent of the curvature of the individual's utility function provided that it is concave and that her intertemporal preference is time separable. Section 6.4.1 discusses the case of non-time-separable preferences where the contract as characterized by Ghili et al. (2022) is no longer optimal.

premium guarantee in the market, the optimal contract offers a constant *consumption guarantee* that *increases* when individuals could get a better consumption guarantee in the market.[13] Moreover, comparing equations (1) with (2), we find that under flat incomes over the lifecycle ($y_t = \bar{y}\ \forall t$), the premium guarantees under the GLTHI result in the optimal consumption guarantees (which, in the case of flat income over the lifecycle, are independent of the income level). We highlight this important result in the following Remark:

Remark 3 *In the special case of flat income profiles over the lifecycle, the GLTHI contract coincides with the optimal contract.*

3.3 Arrow Securities Implementation

Krueger and Uhlig (2006) note that the optimal dynamic long-term contract with one-sided commitment can be implemented through the trade of one-period state-contingent Arrow securities. As such, long-term contracts can be turned into a sequence of short-term contracts. A main benefit of this alternative implementation is that it may enable competition between insurers rather than exposing the individual to the potential monopoly power of the insurer created by the lock-in effect embedded in long-term contracts. A main drawback is the added complexity relative to guaranteed premium (or consumption) profiles under long-term contracts.

The following Lemma characterizes the quantity of securities traded to implement the GLTHI premium path:

Lemma 1 *The premium path of the GLTHI contract can be replicated by purchasing actuarially fair short-term insurance contracts supplemented by Arrow securities. The quantity of Arrow securities bought after history $\Xi_t \equiv (\xi_1, \xi_2, ..., \xi_t)$ that pay one dollar in state ξ_{t+1} is equal to*

$$b_t\left(\xi_{t+1} \mid \Xi_t\right) = \begin{cases} 0 \quad \text{if } P_{t+1}\left(\xi_{t+1}\right) < \tilde{P}_t\left(\Xi_t\right) \\ E(m_{t+1}|\xi_{t+1}) - \tilde{P}_t(\Xi_t) + \\ \sum_{\tau>t+1}^{T} \sum_{z\in\mathcal{Z}} \delta^{\tau-(t+1)}[E(m_\tau|z_\tau) - \tilde{P}_t(\Xi_t)]q_\tau\left(z \mid \xi_{t+1}, \mathbf{P}_{t+2}^{\tau}, \tilde{P}_t(\Xi_t)\right) \quad \text{otherwise,} \end{cases} \tag{3}$$

where (a) $\tilde{P}_t(\Xi_t)$ is the GLTHI premium in period t after history Ξ_t, and (b) $q_\tau\left(z \mid \xi_{t+1}, \mathbf{P}_{t+2}^{\tau}, \tilde{P}_t(\Xi_t)\right)$ is the probability that (i) $\xi_\tau = z$, and (ii) the policyholder does not lapse (or die) between periods $t+1$ and τ, given

[13] We note that guaranteed premium in the German system decreases when consumers switch to a contract with a lower premium guarantee. Although Ghili et al. (2022) derive the optimal contract imposing "no-laspation" contraints, they also note that the same outcome could be achieved by contracts that specify a guaranteed "premium path" that provides a constant consumption floor as long as individuals decide to renew. As such, equivalent to GLTHI, increases in the consumption guarantee under the optimal contract would occur when individuals lapse and sign a contract with a new insurer that offers a higher consumption guarantee (via a lower premium).

the subsequent equilibrium premiums $\mathbf{P}^{\tau}_{t+2}$ *and* $\tilde{P}_t(\Xi_t)$.

Proof 1 *See Appendix A3.*

Lemma 1 states that the Arrow securities to replicate the GLTHI premiums pay either nothing (in the event of lapsing), or the net present value of the difference between the guaranteed-renewable GLTHI premium and the premium of a short-term contract (in the event of not lapsing). Appendix A3.1 provides an analogous expression for the Arrow securities that replicate the optimal dynamic contract in Ghili et al. (2022). Appendix A7 makes use of Lemma 1 and our data to calculate the quantity of securities needed to replicate the GLTHI contract and the optimal dynamic contracts.

4 Claims and Survey Panel Data from Germany

This section describes the GLTHI claims panel data and the SOEP survey panel data. We use the claims data to estimate individual health transitions and expenditures over the lifecycle. We use the SOEP data to estimate individual income dynamics over the lifecycle. The evolution of health expenditures and income paths over the life-cycle are the key inputs to assess welfare under the different health insurance contracts with analyze.[14]

4.1 GLTHI Claims Panel Data

The claims panel data contain the universe of GLTHI contracts and claims between 2005 and 2011 from one of the largest German private health insurers. In total, our data include more than 2.6 million enrollee-year observations from 620 thousand unique policyholders along with detailed information on plan parameters such as premiums, claims, and diagnoses. The data also contain the *age* and *gender* of all policyholders, their occupational group, and the age when they first signed a contract with the insurer. We converted all monetary values to 2016 U.S. dollars (USD). Atal et al. (2019) provide more details about the dataset.

Sample Selection. We focus on primary policyholders. In other words, we disregard children and those below 25 years (555,690 enrollee-year observations).[15] Moreover, due to the 2009 portability reform (see footnote A58), we disregard inflows after 2008 (253,325 enrollee-year observations). Our final sample consists of 1,867,465 enrollee-year observations from 362,783 individuals.

[14]In one extension in Section 6.4.1, we employ a claims dataset from one of the biggest German *public* insurers to assess the robustness of our results to using alternative initial health distributions. Further, we use the Panel Study of Income Dynamics (PSID) survey to assess the robustness of our results to using lifecycle income paths for the United States.

[15]The GLTHI market only features individual policies, not family policies; even children have their individual policy. However, if parents purchase the policy for their child within two months of the birth, no risk-rating applies for the child. Under the age of 21, insurers do not have to budget and charge for old-age provisions.

Descriptive Statistics. Table A1 (Appendix) shows that the mean age of the sample is 45.5 years and the oldest enrollee is 99 years old. Thirty-four percent of the sample are high-income employees, 49 percent are self-employed, and 13 percent are civil servants. The majority of policyholders (72 percent) are male, because women are underrepresented among the self-employed and high-income earners in Germany.

On average, policyholders have been clients of the insurer for 13 years. They have been enrolled in their current health plan for 7 years. Ten percent of all policyholders are clients for more than 28 years; one policyholder even for as long as 86 years, illustrating the existence of a real-world private long-term health insurance system.[16] Figure A2 shows the distribution of policyholders' age at contract inception. The majority of individuals sign their first GLTHI contract at around 30 years old, an age when most Germans have fully entered the labor market but are still healthy and face affordable premiums.

The average *annual premium* is $4,749 and slightly lower than the average premium for a single plan in the U.S. group market at the time (Kaiser Family Foundation, 2019). Note that the *annual premium* is the total premium—including employer contributions for privately insured high-income earners.[17] The average *deductible* is $675 per year.

In terms of benefits covered, we simplify the rich data and focus on three plan-generosity indicators provided by the insurer. These classify plans into *TOP*, *PLUS*, and *ECO* plans. *ECO* plans lack coverage for services such as single rooms in hospitals and treatments by a leading senior M.D. (*Chefarztbehandlung*) that *TOP* and *PLUS* plans offer. For *ECO* and *PLUS* plans, a 20 percent coinsurance rate applies if enrollees see a specialist without referral from their primary care physician, while such coinsurance does not apply for *TOP* plans. About 38 percent of all policyholders have a *TOP* plan, 34 percent a *PLUS* plan, and 29 percent an *ECO* plan. Because these plan characteristics have mechanical effects on claim sizes and correlate with policyholders' age, we control for them when modeling health transitions in Section 5.

4.2 Socio-Economic Panel Study

The German Socio-Economic Panel Study (SOEP) is a representative longitudinal survey. Since 1984, it annually surveys about 10,000 households and 20,0000 individuals above the age of 17 (Goebel et al., 2019). We use SOEPlong and all existing waves from 1984 to 2016, to fully exploit the lifecycle dimension of this panel (SOEP, 2018).

[16]Our insurer doubled the number of clients between the 1980s and 1990s and has thus a relatively young enrollee population, compared to all GLTHI enrollees. Gotthold and Gräber (2015) report that a quarter of all GLTHI enrollees are either retirees or pensioners.

[17]Employers cover roughly one half of the total premium and the self-employed pay the full premium.

Our main income measure is *equivalized post-tax post-transfer annual income*. It accounts for within-household redistribution and controls for economies of scale by assigning each individual a needs-adjusted income measure. Specially, *equivalized, post-tax post-transfer annual income* sums over all post-tax monetary income flows at the household level, such as income from labor, capital, public and private retirement accounts, or social insurance programs.[18] Then, the total annual post-tax household income is divided by the number of household members using the modified OECD equivalence scale.[19]

Sample Selection. We leave the representative sample as unrestricted as possible, but exclude observations with missings on core variables such as age, gender, employment, or the insurance status. Other than that, we only exclude respondents below the age of 25.

Descriptive Statistics. Table A2 (Appendix) provides summary statistics for our SOEP sample. From 1984 to 2016, the average annual income per household member was $26,433 (in 2016 USD). Note that this measure has positive values for *all* respondents.

In the SOEP sample, the average age is 47, and 52 percent are female. About 27 percent are white collar workers, six percent are self-employed, and four percent are civil servants. Forty-two percent work full-time and 14 percent part-time.

5 Modeling Health Risks and Income over the Lifecycle

5.1 Risk Classification

Risk classification is a key ingredient for calculating the premiums and welfare of short- and long-term insurance contracts. We rely on insights from actuarial science to produce an "efficient" risk classification. Our method aims to improve upon the state-of-the-art literature by better representing skewed health expenditure distributions. We present below the features of our method and relegate the details to Appendix A5.

Following the economics literature (Einav et al., 2013; Handel et al., 2015; Ghili et al., 2022), we begin by constructing an individual measure for expected health care costs using the German version of the John Hopkins ACG software, which is routinely used by insurers for underwriting. For each individual, the ACG software provides a *continuous risk score* λ_t^* which represents expected health

[18] The SOEP group also generates and provides these single components in a time-consistent manner.

[19] The modified OECD equivalence scale assigns a value of 1 to the household head, 0.5 to other adults, and 0.3 to children up to 14 years of age.

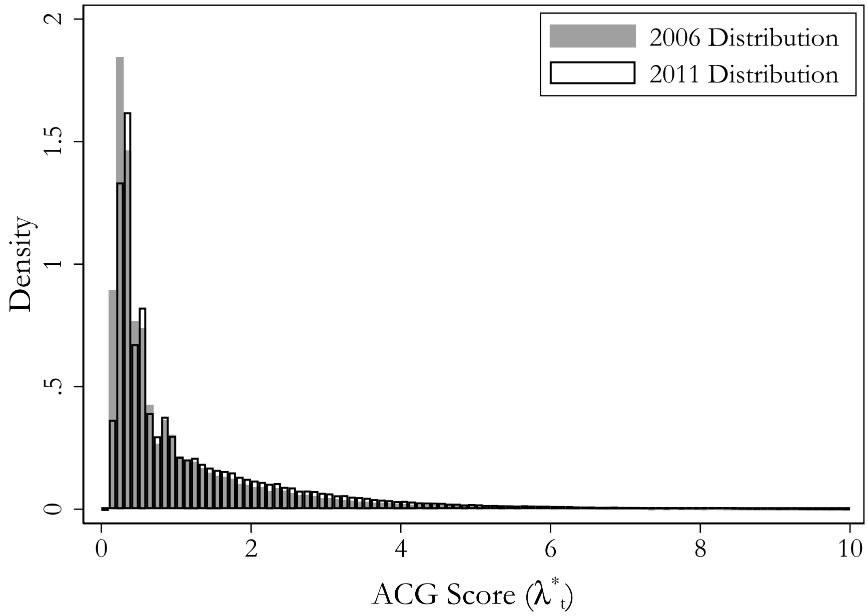


Figure 1: Distribution of λ_t^* in 2006 and 2011

Notes: The figue shows the distribution of ACG scores for years 2006 and 2011. The distribution of λ_t^* is truncated at 10; 0.7 percent of the sample have $\lambda_t^* > 10$. Source: Own calculations using GLTHI claims data and ACG Software.

care costs in year t relative to the mean in the reference population.[20] It is based on (a) diagnosis codes (pre-existing conditions and claim diagnoses), (b) costs of treatments, (c) treatment episode dates, and (d) age and gender.

Figure 1 shows the empirical distributions of λ_t^* for our GLTHI claims data in 2006 (the first year in our sample) and 2011 (the last year in our sample). Both distributions are approximately unimodal, and appear stable over time.[21] It also illustrates that the distribution of λ_t^* is heavily skewed and has a long right tail. For example, the top percentile of λ^* has expected health expenditures $\mathbb{E}\left(m|\lambda^* \geq \mathbb{P}_{99}\right)$= \$63,422, the second highest percentile has $\mathbb{E}\left(m|\mathbb{P}_{98} \leq \lambda^* < \mathbb{P}_{99}\right) = \$30,027$, and the following three percentiles have $\mathbb{E}\left(m|\mathbb{P}_{95} \leq \lambda^* < \mathbb{P}_{98}\right) = \$19,253$, where $\mathbb{P}_k$ denotes the k-th percentile of the distribution in Figure 1. The distribution of expected expenditures thus mirrors the characteristic long right tail of health expenditure distributions (French and Kelly, 2016).

Our method combines the continuous score λ_t^* and its $n-1$ lags into the vector of scores, $\Lambda_t^*(n) \equiv \{\lambda_t^*, \lambda_{t-1}^*, .., \lambda_{t-n-1}^*\}$, which we then map into K different *discrete risk categories*. Modeling risk types with discrete categories serves two specific purposes. First, as we allow the contract premiums to depend on the risk type, the granularity in our model should capture the granularity of the information used by the underwriters, both in the actual environment and in counterfactual scenarios. Second, the model should be parsimonious enough to allow for modeling health dynamics with a reasonable

[20]In this case, the reference population are the publicly insured individuals in Germany

[21]This also suggests that excluding inflows of new enrollees in 2010 and 2011 due to the portability reform (see Section 2) poses no major issue.

number of parameters. The skewness in Figure 1 implies that the amount of reclassification risk will strongly depend on the granularity of the risk classification.

Our main methodological contribution to risk classification modelling is in how to discretize health risk. The commonly-used approach would use an ad-hoc criterion to partition λ_t^* into different discrete categories. We depart from the common approach in two key ways: First, we allow health status to be a function of current and lagged values of λ_t^*, i.e., $\Lambda_t^*(n)$, where we determine n within our procedure. Our approach can therefore allow for higher-order dependencies in the health dynamics in a parsimonious way. Second, we propose and implement a method to discretize the vector of scores $\Lambda_t^*(n)$ into an endogenously determined number of risk categories K and corresponding partitions. The method maximizes an efficiency criterion from the actuarial science literature (cf. Finger, 2001).

We split the task of constructing the discrete risk categories $\{\lambda_1, .., \lambda_K\}$ into two sequential problems: (1) For a given number of categories K, define the *efficient partition* of the scores vector $\Lambda_t^*(n)$ into K discrete categories. (2) Select the parameters K and n that lead to the best performance of the classification system. We expand on each step below.

Efficient Partitioning. According to the actuarial science literature (Finger, 2001), an efficient risk classification system has two properties: *homogeneity*—meaning that individuals in the same risk category have similar risk; and *separation*—meaning that the categories have sufficiently different expected claims to justify distinct categories. For instance, in Figure 1, equally-sized categories are unlikely to be optimal as they would assign similar individuals in terms of λ^* into different categories in the left tail of the distribution, failing the *separation* principle. In addition, it would assign individuals with substantial λ^* differences into identical categories in the right tail of the distribution, failing the *homogeneity* principle. Appendix A5 shows that applying k-means clustering to $\mathbb{E}(m_t \mid \Lambda_t^*(n))$—i.e., to the mean claims by an individual's ACG score—yields an efficient classification.

Parameter Selection. Next, we determine the number of lags n of ACG scores when computing $\mathbb{E}(m_t \mid \Lambda_t^*(n))$ and the number of risk categories K.[22] k-means clustering is an unsupervised learning method; we assume that the overall objective applies also when selecting K and n. As shown in Appendix A5, this implies using the R^2 of a regression of expenditures on risk category indicators as our criterion for model selection.

Figure 2 shows how the performance depends on K and n. For all values of n, initially, the

[22]Including lagged ACG scores is consistent with an underwriting process often covering a relatively long medical history of the applicant (e.g., all diseases of the past 5 years and all surgeries of the past 10 years in case of our insurer).

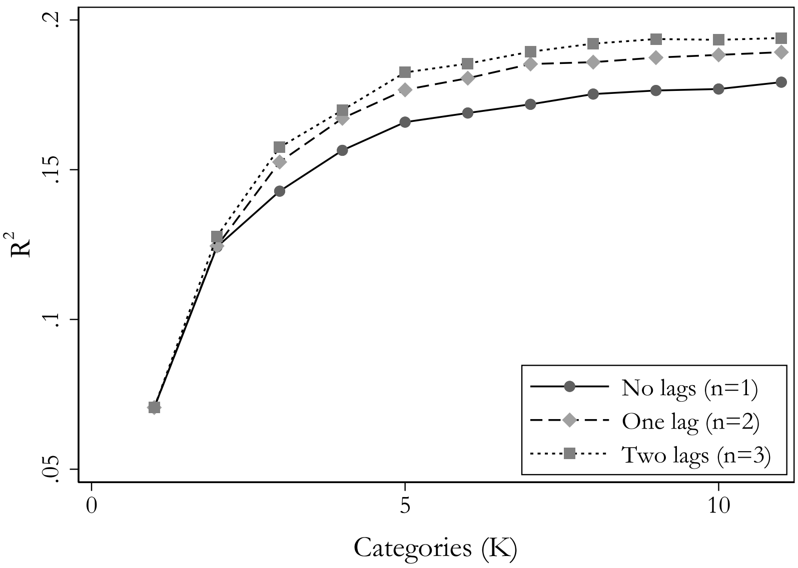


Figure 2: Performance of Alternative Risk Classifications.

Note: The figure displays *unadjusted* R^2 of a regression of expenditures on risk category indicators. All results are robust to using the *adjusted* R^2. Each specification includes 21 age-gender fixed effects, year fixed effects and 79 plan fixed effects. Source: Own Calculations using German Claims Panel Data.

predictive power improves rapidly when we increase the number of categories K; however, this improvement levels out at around $K = 5$. Moreover, compared to only using the previous year's score ($n = 1$), there is distinct improvement when adding one lag ($n = 2$) whenever $K \geq 3$. However, adding a second lag only marginally improves the predictive accuracy. Figure 2 shows that, beyond including one lag ($n = 2$) and 7 distinct risk categories, increasing K or n further yields negligible improvement in performance. Thus as a compromise of predictive accuracy and model parsimony, we choose $K = 7$ and $n = 2$ (in the spirit of Heckman and Burton Singer (1984) and Keane and Wolpin (1997) in their choice of the number of unobserved types in the labor economics literature).[23]

5.2 Estimation of Transition Matrices and Mean Expenditures

We posit that the health status of individual i at age t, $\xi_{it} \equiv (A_{it}, \lambda_{it})$, depends on contemporaneous risk λ_{it} and age A_{it}, where A_{it} is one of eleven age groups (five-year bands from age 25 to age 75 and 75+).[24]

As our risk classification generates risk categories of very different sizes, we use a parametric, yet flexible, model for transition rates between discrete risk categories λ_t and for mean expenditures

[23]Section A5 provides additional robustness checks, including the role of outliers, including a longer history of claims, and sample changes when n changes. We also show that transition rates between risk categories satisfy first-order stochastic dominance as assumed in Ghili et al. (2022).

[24]Note that the ACG scores are based on an individual's age, so that, in principle, a risk category λ_{it} that uses ACG scores as input should contain all the information needed to predict mean expenditures. However, ACG scores are not designed to predict transitions so, in principle, transition matrices may depend on age even after conditioning on λ_{it}. As discussed below, our results confirm these predictions.

by risk category and age. Specifically, to estimate the transition matrices for health dynamics, we estimate a multinomial logit model:

$$\eta_{it}^{j} = A_{it}\beta_j + L_{it}\gamma_j + h\left(A_{it}, L_{it}; \theta_j\right) + \epsilon_{it}^{j}, \tag{4}$$

where η_{it}^{j} represents the log odds for $\lambda_{it+1} = j$, for $j \in \{2, \ldots, 8\}$ relative to the reference category $\lambda_{it+1} = 1$ ($\lambda_{it+1} = 8$ represents death); A_{it} represents i's age groups, and L_{it} is a set of indicators for the categories of λ_{it}. $h\left(A_{it}, L_{it}; \theta_j\right)$ consists of pairwise interactions of A_{it} and L_{it} with the associated parameter vector θ_j.[25]

To model the expected claims based on risk category, we use predicted values from an OLS regression. In addition to the controls in Equation (4), we also control for Q_{it} representing health plan generosity dummies $q \in \{ECO, PLUS, TOP\}$. The base specification is:

$$m_{it} = A_{it}\beta + L_{it}\gamma + Q_{it}\delta + h\left(A_{it}, L_{it}, Q_{it}; \theta\right) + \epsilon_{it}. \tag{5}$$

In an iterative process, we add pairwise interaction terms between A_{it}, L_{it}, and Q_{it} (represented by $h\left(A_{it}, L_{it}, Q_{it}; \theta\right)$) to Equation (5) until no remaining term is statistically significant.[26]

Descriptive Statistics for Transition Matrices. Table 1 displays one-year transition rates between health risk categories for all age groups; the numbers are predicted probabilities based on Equation (4). Two facts emerge from Table 1. First, we find strong persistence in health risk. For instance, an individual with $\lambda_t = 1$ has an 83 percent probability of $\lambda_{t+1} = 1$. The likelihood of staying in the same category between two consecutive years generally decreases over risk categories but, still, forty-five percent of individuals in category 7 remain in category 7 in the next year. Second, despite the high persistence, the likelihood of reclassification is non-trivial even when just considering two subsequent years. For example, the probability of ending up in a different risk category in $t+1$ (with very different mean expenditures as shown below) is more than 50% for category 2 in year t.

The transition rates are highly dependent on age. Tables A3 and A4 (Appendix) show the transition matrices separately for each of the 11 age groups. For example, the probability of remaining in category 1 decreases from 89 percent among 25-year-olds to 18 percent among individuals above 75. Also the probability of recovering, i.e. transitioning from a higher to a lower risk category, declines

[25]We select the interaction terms sequentially: in each iteration, we include the interaction term with the strongest association with transition rates (based on a χ^2 test), until none of the remaining interaction terms is statistically significant.

[26]We use a subsample of policyholders with moderately-sized deductibles to estimate conditional expenditures given λ_t as policyholders with large deductibles may decide not to submit their claims, leading to downward biased estimates. In Appendix A5 we provide some descriptive statistics for this subsample, which generally confirm that this assumption is reasonable.

Table 1: Health Risk Category Transitions

	λ_{t+1}							
λ_t	1	2	3	4	5	6	7	8 (†)
1	0.831	0.158	0.006	0.003	0.001	0.001	0.000	0.001
2	0.214	0.523	0.215	0.036	0.009	0.001	0.000	0.002
3	0.050	0.179	0.572	0.164	0.029	0.003	0.000	0.003
4	0.024	0.053	0.227	0.541	0.128	0.013	0.001	0.013
5	0.018	0.027	0.035	0.330	0.445	0.104	0.005	0.036
6	0.010	0.018	0.017	0.096	0.294	0.409	0.052	0.104
7	0.002	0.005	0.002	0.027	0.085	0.200	0.452	0.226

Notes: The table shows average transition probabilities across the different health categories. The sample includes all years, all age groups, and uses the ACG scores to construct risk categories λ as explained in Section 5.1. Source: Own calculations based on German Claims Panel Data.

with age. Moreover, the mortality rates increase rapidly with age—in particular for states below 7. All these differences are statistically significant. Therefore, allowing for age-dependent transition rates is necessary.

Descriptive Statistics for Expenditures. Table 2 shows mean expenditures m by age group, and the distribution of risk categories within each group. As expected, mean expenditures strongly increase in age: they almost double from \$1,996 in the age group 25 to 30, to \$3,719 in the age group 45 to 50, and almost double again to \$7,151 in the age group 65 to 70. For those above 75 years, the average is \$10,020 (all values in 2016 U.S. dollars). This age gradient is, however, accounted for by our risk classification.[27]

There is also a clear age gradient in health risk. The probability of $\lambda = 1$ declines progressively with age, whereas the share of enrollees in the five highest categories increases with age. Only 1.7 percent of those in the age group 25 to 30 are in categories $\lambda = 4$ and $\lambda = 5$. This share almost quadruples to 6.2 percent for the age group 45 to 50, and then more than quadruples again to 28.6 percent for the age group 65 to 70. It is 61 percent for enrollees above 75 years. Risk category $\lambda = 7$ clearly represents catastrophic risk and covers at most 0.3 percent in any age group in a given year. Section 6.5 discusses how catastrophic risk affects welfare under the different contracts.

[27] As the ACG score depends on age, age-group indicators A_{it} should not have predictive power in the model for expected expenditures if our risk classification based on ACG scores is rich and flexible enough. Even though a few age-related parameters in Equation (5) turn out statistically significant, the deviations from mean expenditure within each risk category are economically insignificant. Figure A6 (in Appendix A5) illustrates this point. We interpret it as evidence that our preferred risk classification is rich enough. Therefore, we restrict all age effects to zero.

Table 2: Health Expenditures and Risk Categories λ by Age Group

Age	Mean	S.D.($\mathbb{E}(m \mid \lambda)$)	λ_t						
			1 (Healthiest)	2	3	4	5	6	7 (Sickest)
25-30	1,996	1,782	0.789	0.154	0.039	0.013	0.004	0.001	0.000
30-35	2,619	1,938	0.740	0.178	0.054	0.020	0.006	0.001	0.000
35-40	2,840	2,086	0.652	0.225	0.085	0.027	0.009	0.002	0.000
40-45	3,119	2,411	0.622	0.227	0.103	0.034	0.012	0.003	0.000
45-50	3,719	2,946	0.539	0.258	0.136	0.046	0.016	0.004	0.001
50-55	4,880	3,544	0.463	0.263	0.174	0.068	0.024	0.007	0.001
55-60	6,517	4,573	0.291	0.319	0.232	0.108	0.036	0.011	0.002
60-65	7,635	4,299	0.184	0.313	0.269	0.155	0.058	0.019	0.003
65-70	7,151	4,421	0.069	0.291	0.337	0.217	0.069	0.014	0.002
70-75	8,355	5,026	0.019	0.203	0.347	0.309	0.105	0.015	0.002
75+	10,020	4,490	0.000	0.092	0.267	0.422	0.188	0.029	0.003

Notes: The table shows average claims and the standard deviation of claims within each risk category and age group for the different health categories. It also shows the fraction of individuals in each category by age group. The sample includes all years, all age groups, and uses the ACG scores to construct risk categories λ as explained in Section 5.1. Source: Own calculations based on German Claims Panel Data.

5.3 Lifecycle Income Paths

Next, we estimate the lifecycle income paths using 33 years of the German Socioeconomic Panel Survey (SOEP) (1984-2016). We estimate income over the entire lifecycle as we are interested in assessing the performance of long-term contracts that are guaranteed renewable until death. Our income measure is *equivalized post-tax post-transfer annual income*, which sums over all post-tax income flows at the household level, and then normalizes by the number of household members. We estimate the following individual fixed effects model:

$$log(y_{it}) = \theta_i + f(age_{it}) + \epsilon_{it}, \tag{6}$$

where y_{it} is our income measure in 2016 U.S. dollars for individual i in year t; and θ_i are individual fixed effects which net out persistent individual time-invariant income determinants, such as gender, preferences, or work productivity. The flexible function $f(age_{it})$ represents a series of age fixed effects and identifies the main coefficients of interest.

Because lifecycle profiles differ substantially by educational degree (Becker and Chiswick, 1966; Bhuller et al., 2017), we estimate the income process by education for (a) individuals with the highest schooling degree after 13 years of schooling (*Ed 13*), and (b) individuals with an intermediate degree after 10 years of schooling (*Ed 10*).[28]

The two dashed curves in Figure 3 show the estimated age fixed effects for people with 10 years

[28] Germany has three different schooling tracks where the majority of students complete school after 10 years and then start a three-year apprenticeship (cf. Dustmann et al., 2017).

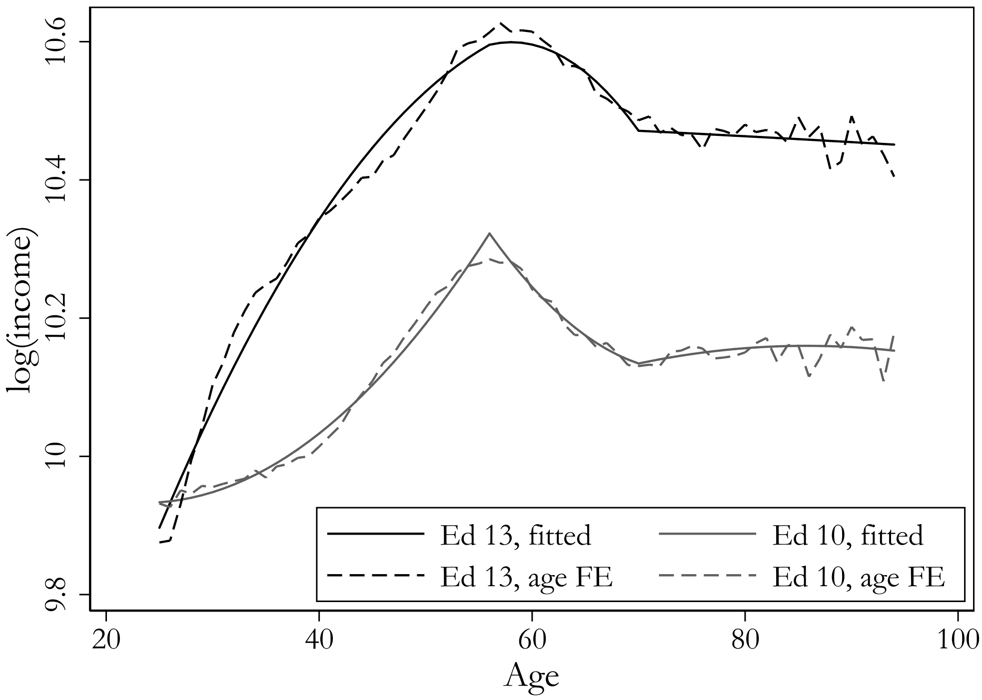


Figure 3: Life-cycle Income Paths Germany, Nonparametric and Fitted.

Notes: The figure shows the predicted income using Equation (6), and the fitted, piece-wise squared polynomial. All values in 2016 USD. Source: Own calculations using SOEP (2018) for the years 1984 to 2016.

and 13 years of schooling, respectively. Income rises sharply between age 25 and age 57. Then it decreases substantially until around age 70, from which point it remains relatively flat until death. There exists a level difference in income paths between the two educational groups over the entire lifecycle. We accommodate these patterns by fitting $f(age_{it})$ as a piece-wise squared polynomial of age, where we allow the parameters of age and age^2 to differ by education and across three different age bins: $[25, 56]$, $[56, 70]$ and $70+$. This is illustrated by the two solid lines in Figure 3. Note that the piece-wise squared polynomials fit the empirical lifecycle profiles very well.

Several factors can explain these lifecycle income patterns. First, the labor market entry and subsequent careers significantly increase post-tax income between the main working ages 25 and 55. Second, our income measure includes social insurance benefits, and the German welfare state is known for its generosity. Third, equivalized household income starts to decrease after age 57 until around age 70. Especially in the 1980s and 1990s but also today, many Germans retire early (Börsch-Supan and Jürges, 2012); others reduce their working hours, for example, to take care of their grandchildren or provide long-term care for their parents (Schmitz and Westphal, 2017). Finally, the stable permanent income stream from age 70 until death may be explained by the fact that our income measure includes primarily statutory pensions, employer-based pensions and private pensions (Geyer and Steiner, 2014; Kluth and Gasche, 2016).

6 Main Results

6.1 Equilibrium Lifecycle GLTHI Premiums

After estimating the health risk process, we can calculate the equilibrium GLTHI lifecycle premiums by solving Equation (1) using backwards induction. We use a discount factor $\delta = 0.966$, which corresponds to a discount rate of 3.5 percent. Note that $P_t(\xi_t)$ in Equation (1) is the guaranteed-renewable premium that an individual with health status ξ_t would be offered if she entered a contract in period t in the GLTHI market. Therefore, the equilibrium GLTHI premiums correspond to 490 values: premiums depend on an individual's current risk category $\lambda_t \in \{1, 2, ..., 7\}$, as well as age $t \in \{25, ..., 94\}$.

Figure 4 plots the resulting premiums calculated using Equation (1) for a handful of the most relevant combinations: $\lambda_t = 1$ and $t \in \{25, ..., 59\}$; $\lambda_t = 2$ and $t \in \{25, ..., 74\}$; $\lambda_t = 3$ and $t \in \{25, ..., 94\}$; $\lambda_t = 4$ and $t \in \{60, ..., 74\}$; $\lambda_t = 5$ and $t \in \{75, ..., 94]\}$. These combinations represent the most common health categories for each corresponding age interval.

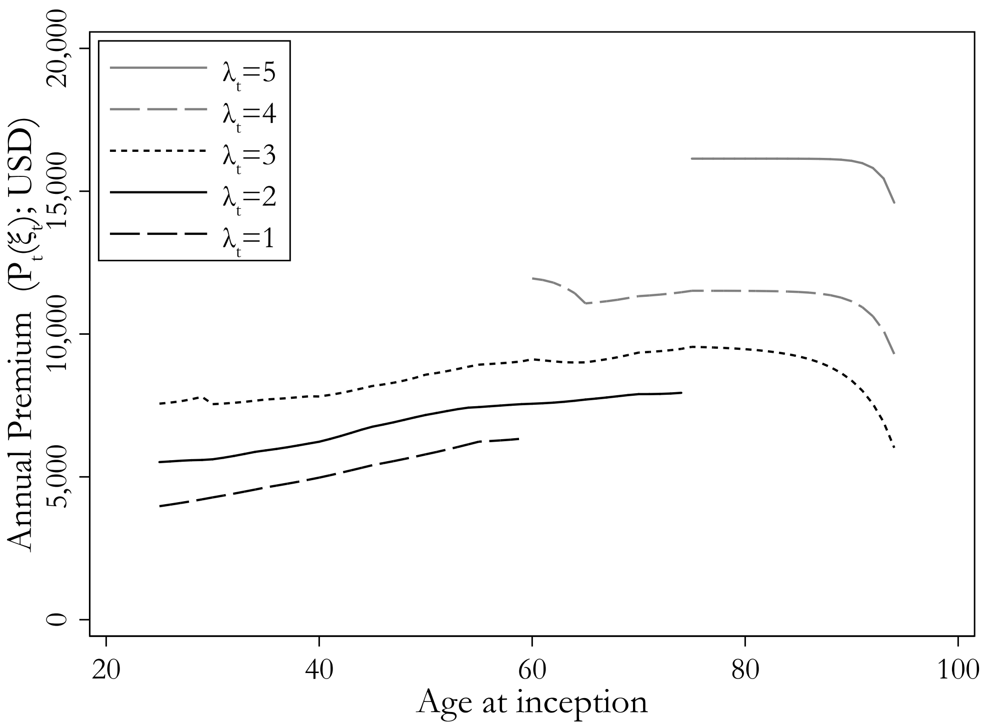


Figure 4: Calibrated Starting Premiums $P_t(\xi_t)$ in the GLTHI.

Notes: Calibrated GLTHI premiums at inception, calculated using Equation (1) and the estimated health status process. Health status ξ is comprised of age and health category λ, as explained in the text. Source: Own calculations using German Claims Panel Data.

Three forces determine the lifecycle profile of $P_t(\xi_t)$ in Figure 4. First, $P_t(\xi_t)$ is an increasing function of ξ_t because for any age, a higher risk category is associated with higher current and future health care claims.

Second, starting premiums increase with age for most age ranges. This is because health transitions depend strongly on age (through the A_t component of ξ_t, see Equation (4)). As a consequence, the annualized net present value of health care claims of an individual with a given λ_t increases with age for most of the age ranges.

Third, starting premiums decrease with age in very old age ranges. The reason is as follows: individuals who renew their contract are an adversely selected subset of contract holders, i.e., those who either remain or become sick enough to not get better outside offers in the market. By charging a front-loaded premium that considers this dynamic adverse selection, the insurer breaks even. However, for any given risk category, the probability of transitioning towards a worse category in the future decreases with age. Therefore, the need to front-load premiums to fund future negative health shocks *decreases* over the lifecycle. This force explains why $P_t(\xi_t)$ decreases with t when t is sufficiently large.

Appendix A6 compares the calibrated and the observed premiums by age at inception for ages between 25 and 75.[29] Both are very similar in the key aspects that determine welfare over the life-cycle. Calibrated premiums for 25 years-old starting in categories 1 and 2 (99.35 percent of the sample) are very similar to the observed ones (and somewhat smaller than observed premiums for sicker individuals). We also observe positively sloped starting premiums by age over this entire age range, both for the calibrated and the observed premiums. Also, the rank ordering of premiums by health status persists over the entire life-cycle.[30]

6.2 Comparison between the GLTHI and the Optimal Dynamic Contract

This subsection compares the lifecycle premiums and the amount of front-loading between the flat-income-optimal GLTHI design and the optimal dynamic (GHHW) contract (Ghili et al., 2022). Using our empirical health status transitions and income dynamics, Table 3 illustrates the differences between the flat-income-optimal GLTHI and the optimal contract by comparing the contract terms at age 25. Panel (a) shows the GLTHI premium and front-loading amounts for a 25-year-old by the

[29]For those health categories with enough observations of new policies being issued.

[30]Although we could, in principle, use the observed premiums to evaluate welfare under the GLTHI, there are two reasons to use the calibrated premiums instead: First, our analysis of the market equilibrium under different assumptions requires knowing the premiums for all possible combinations of ξ and t. Many of these are either completely absent or represented by only a small number of individuals in our data. Second, as we compare the welfare properties of GLTHI to counterfactual, model-based, contracts, we prefer to conduct welfare comparisons based on premiums that were generated in an analogous manner. Note also that the similarities between the observed and theoretical premiums highlighted in the text imply that the welfare under the GLTHI contracts evaluated with the observed premiums should be very close to our results derived from the calibrated premiums. Calibrated starting premiums for the vast majority of individuals are very similar to the observed ones and most people 'lock-in' this premium for their entire life-cycle. Also, note that lapsation rates depend on the rank-order of premiums but not the magnitude of the premium differences.

health risk category $\lambda_{25} \in \{1, ..., 7\}$. If $\lambda_{25} = 1$, she pays a premium of \$3,973, which is \$2,499 in excess of expected claims. Individuals with higher λ's pay higher premiums, but the amount of front-loading decreases. For example, for $\lambda_{25} = 3$ the premium is \$7,563 which includes \$1,545 in front-loading. The reason is that when the current health status decreases, the likelihood of a further health deterioration also decreases.

Table 3: Comparing GLTHI Contract to Optimal Contracts Terms at Inception, Age 25

λ_{25}	1	2	3	4	5	6	7
Expected claims	1,473	3,559	6,019	9,302	14,600	24,554	54,930
(a) GLTHI							
Premium	3,973	5,517	7,563	10,363	15,291	24,561	54,930
Front-loading	2,499	1,957	1,545	1,062	691	7	0
(b) Optimal contract *Ed 10*							
Premium	2,571	5,366	7,489	10,307	15,273	24,554	54,930
Front-loading	1,097	1,807	1,471	1,006	673	0	0
(c) Optimal contract *Ed 13*							
Premium	1,895	4,578	6,988	10,103	15,187	24,554	54,930
Front-loading	421	1,019	970	801	586	0	0

Notes: The table shows expected health care claims, starting premiums, and the amount of front-loading by health risk category at age 25, $\lambda_{25} \in \{1, ..., 7\}$. All values are in 2016 USD. Source: German Claims Panel Data, SOEP (2018).

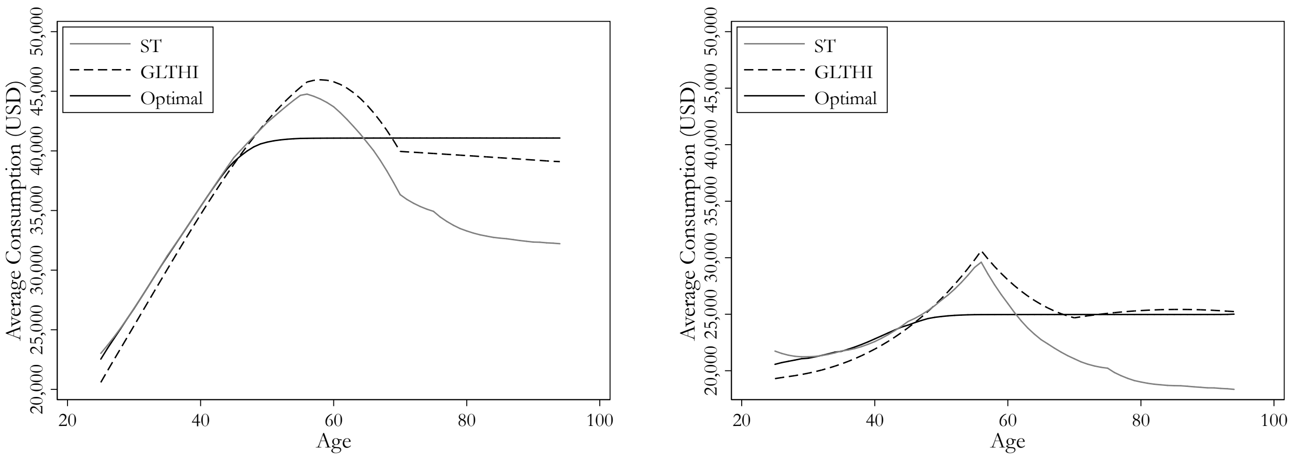
Panel (b) of Table 3 shows the premiums and front-loading amounts for the optimal dynamic contract and an individual with 10 years of schooling (*Ed 10*) by initial health at age 25. For all health categories expect the worst, compared to GLTHI, the initial premiums and front-loading amounts are lower in the optimal dynamic contract. The optimal contract entails less front-loading than GLTHI because of the higher marginal utility of consumption at early ages. However, the differences in premiums between the GLTHI and the optimal dynamic contract decrease as the health status at contract inception worsens. For $\lambda_{25} = 1$ the optimal premium is \$2,571 (vs. \$3,973 for GLTHI) and for $\lambda_{25} = 4$, the optimal premium is \$10,307 (vs. \$10,363 for GLTHI). Note, however, in contrast to the constant guaranteed premiums under the GLTHI, the premiums under the optimal contracts need to adjust as income evolves to maintain any given level of consumption guarantee.

Panel (c) of Table 3 shows the premiums and front-loading amounts for the optimal dynamic contract and an individual with 13 years of schooling (*Ed 13*) by initial health at age 25.This individual has a steeper income profile over her lifecycle (see Figure 3), which is why the optimal contract entails a *lower* degree of front-loading than for *Ed 10*, especially for healthy individuals. Again, the sicker the individual is at inception, the lower the front-loading amount.

Equivalent Arrow Securities. As discussed in Section 3.3, it is possible to replicate the dynamic contracts with one-period Arrow securities. Appendix A7 provides numerical examples for the terms of such contracts for the GLTHI contract and the optimal contract. As expected, the quantity of securities needed to replicate the optimal contract is smaller than for the GLTHI contract. This reflects the lower degree of front-loading required under the optimal contract.

6.3 Consumption Profiles over the Life-cycle

Life-cycle Consumption and Intertemporal Consumption Smoothing. We now explore how the short-term contract, the flat-income-optimal GLTHI contract, and the optimal contract affect intertemporal consumption smoothing and reclassification risk over the life-cycle. To do so, we simulate the consumption lice-cycle profiles, from age 25 to age 94, for $N = 500,000$ individuals. Figure 5 plots the *average* consumption for these three contracts over the life-cycle, separately for *Ed 10* (Figure 5a), and *Ed 13* (Figure 5b).



(a) *Ed 13* (b) *Ed 10*

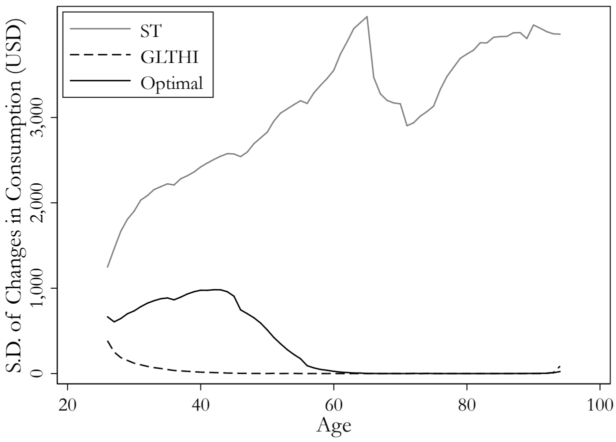
Figure 5: Simulated Average Consumption over the Lifecycle by Education

Notes: The figure displays average simulated consumption under short-term contracts, the GLTHI contract, and the optimal contracts. Panel (a) shows results for the steeper income profile (Ed 13), Panel (b) shows result for the flatter income profile (Ed 10). Source: Own calculations using German Claims Panel Data and SOEP (2018).

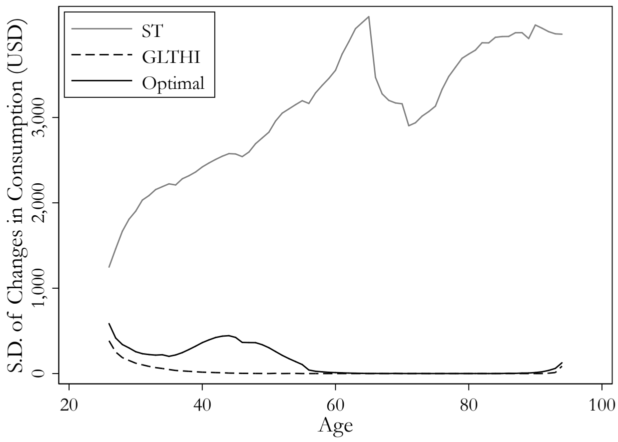
Under a series of short-term contracts, average consumption is simply income minus expected health expenditures. As shown by the grey solid lines, the average consumption profile under short term contracts is therefore hump-shaped over the life-cycle for both education groups. As shown by the dashed lines, under the GLTHI contract, average consumption has a similar shape, but starts at a lower level and is higher at older ages. This reflects the heavy front-loading of GLTHI up to the early 50s. As shown by the black solid lines, compared to GLTHI, the average consumption under the optimal dynamic contract would start at a higher level, particularly for the highly educated who have steeper income profiles and for whom front-loading is costlier.

On top of the lower frontloading in the initial periods, studying income profiles over the life-cycle highlights an additional important advantage of optimal long-term contracts. In our setting with hump-shaped income profiles, the optimal contract allows to fully smooth consumption intertemporally after some age, as the straight flat black consumption line after around age 40 illustrates. The optimal contract prescribes higher premiums in the middle ages, not only to finance reclassification risk insurance, but also to shift resources towards later years when income is low.

Reclassification Risk. To illustrate the degree of *reclassification risk* over the life-cycle, Figure 6 displays the standard deviations of consumption changes over the life-cycle. That is, Figure 6 plots, for each age t, the standard deviation of $\Delta C_{i,t} \equiv C_{i,t+1} - C_{i,t}$ across individuals i.



(a) *Ed 13*



(b) *Ed 10*

Figure 6: Simulated Standard Deviation of Consumption Changes over the Life-cycle, by Education

Notes: The figure displays the simulated standard deviation of consumption changes under short-term contracts, the GLTHI contract, and the optimal contracts. Panel (a) shows results for the steeper income profile (Ed 13), Panel (b) shows result for the flatter income profile (Ed 10). Source: Own calculations using German Claims Panel Data and SOEP (2018).

As seen, the GLTHI contract imposes very little reclassification risk as most individuals lock in $P_{25}(\cdot)$ in the first period. The few individuals who switch contracts are those who start with $\lambda_{25} > 1$ and become sufficiently healthier over time (such that $P_t(\xi_t) < P_{25}(\xi_{25})$ for some $t > 25$). However, this is a rare event, especially after age 40. On the other hand, the optimal dynamic contract entails consumption bumps early in life. For instance, the consumption guarantee under the optimal contract increases for individuals who start at $\lambda_{25} = 1$ and remain at $\lambda_{26} = 1$ in the following year. The reason is that a competing insurer that considers the "good news" regarding future health, contained in the event "$\lambda_{25} = 1$ and $\lambda_{26} = 1$," can offer the individual a higher consumption guarantee, and still break even in expectation. Finally, the standard deviation of consumption changes increases strongly between age 25 and 60 for a series of short-term contracts, then decreases slightly up to age 70, and then increases again until death. Relatedly, Figure A12 (Appendix) shows lapsation rates under each contract.

6.4 Baseline Welfare Results

We now calculate how the differences in consumption profiles over the life-cycle translate into welfare differences across the different contracts. We summarize lifetime utility under each contract with the Consumption Certainty Equivalent (CE). In this section, we focus on the relative welfare difference between the GLTHI contract and the optimal dynamic contracts. We also compare the CE of each long-term contract with welfare under short-term contracts, and with the first-best consumption level, which is equal to the annualized present discounted value of "net income" $y_t - \mathbb{E}(m_t)$, taking into account mortality risk. Details on these welfare measures are in Appendix A9.

So far, we have not specified the utility function because the premiums in the long term contracts do not hinge on a specific utility function. For welfare comparisons, we use a constant absolute risk aversion (CARA) utility function of the form:[31]

$$u(c) = -\frac{1}{\gamma}e^{-\gamma c}. \tag{7}$$

In our main results, we use a risk aversion parameter $\gamma = 0.0004$ (Ghili et al., 2022). In Section 6.4.1, we explore the robustness to (i) the level of risk aversion, (ii) using constant relative risk aversion (CRRA) utility function, and (iii) to using non-time-separable Epstein-Zin preferences, among other robustness tests.

We provide eight sets of results in Table 4, stratifying the findings by the two education groups with different lifecycle income paths. Each corresponds to a different probability simplex that determines the initial category, $\Delta_0 \in \Delta^7$. For instance, Panel (a) assumes that everyone starts in the healthiest category, such that $\Delta_0 = \frac{1}{100}[100, 0, 0, 0, 0, 0, 0]$. Panel (h) shows our baseline results, which assume that λ_{25} is drawn from the distribution implied by the transition matrix at age 25, given $\lambda_{24} = 1$ (see Table A3, Appendix). By doing so, we accurately replicate the distribution of ξ among the 25 to 30 age group. In Panel (h), we also assume that individuals cannot start in the worst health category. This assumption is motivated by the fact that individuals starting in category 7 are unable to afford actuarially fair premiums. One possibility is a fall-back option, e.g. a coexisting public health insurance system.

Column (1) shows welfare under the first-best contract, C^*; Column (2) shows welfare under a series of short-term contracts, C_{ST}. Column (3) shows welfare under the optimal contracts, C_{GHHW}.

[31]The CARA utility function has the convenience of allowing for negative consumption, which occurs when income is lower than the required premium payments. However, it also implies that the consumption equivalent may be negative under some contracts (see Table 4).

Table 4: Welfare under different contracts

	(1) C^*	(2) C_{ST}	(3) C_{GHHW}	(4) C_{GLTHI}	(5) $\frac{C_{GHHW}-C_{GLTHI}}{C_{GHHW}}$
	Panel (a): $\Delta_0 = \frac{1}{100}[100,0,0,0,0,0,0]$				
Ed 10	23,029	-10,156	22,489	21,537	4.2%
Ed 13	34,210	-2,396	27,727	26,025	6.1%
	Panel (b): $\Delta_0 = \frac{1}{100}[0,100,0,0,0,0,0]$				
Ed 10	22,600	-10,815	21,373	20,840	2.5%
Ed 13	33,776	-4,074	25,569	24,898	2.6%
	Panel (c): $\Delta_0 = \frac{1}{100}[0,0,100,0,0,0,0]$				
Ed 10	22,244	-10,704	20,169	19,855	1.6%
Ed 13	33,420	-2,382	23,619	23,273	1.5%
	Panel (d): $\Delta_0 = \frac{1}{100}[0,0,0,100,0,0,0]$				
Ed 10	21,905	-10,844	18,409	18,253	0.8%
Ed 13	33,079	-2,416	21,100	20,945	0.7%
	Panel (e): $\Delta_0 = \frac{1}{100}[0,0,0,0,100,0,0]$				
Ed 10	21,476	-10,970	14,715	14,679	0.2%
Ed 13	32,648	-2,473	16,646	16,598	0.3%
	Panel (f): $\Delta_0 = \frac{1}{100}[0,0,0,0,0,100,0]$				
Ed 10	20,639	-11,224	5,969	5,968	0.0%
Ed 13	31,809	-2,605	7,574	7,568	0.1%
	Panel (g): $\Delta_0 = \frac{1}{100}[0,0,0,0,0,0,100]$				
Ed 10	11,587	-27,085	-27,070	-27,070	0.0%
Ed 13	22,327	-24,631	-24,629	-24,629	0.0%
	Panel (h): $\Delta_0 = \frac{1}{100}[89.10,10.25,0.47,0.11,0.04,0.03,0]$				
Ed 10	22,980	-10,119	21,945	21,168	3.5%
Ed 13	34,159	-2,223	26,093	25,088	3.9%

Notes: The table shows welfare measured by the Consumption Certainty Equivalents in 2016 USD dollars, per capita, per year, separately for two income profiles (see Figure 3). Panels (a) to (g) differentiate by initial health status $\lambda_{25} \in \{1,...,7\}$. In Panel (h), we do not allow 25-year-olds to be in the worst health risk category. Columns (1) to (4) show welfare according to the (1) first-best (C^*), (2) a series of short-term contracts (C_{ST}), (3) the optimal contract (C_{GHHW}), and (4) the GLTHI (C_{GHHW}). Column (5) shows the percentage of welfare loss under GLTHI relative to the optimal contract. German Claims Panel Data, SOEP (2018).

Column (4) shows welfare under the GLTHI contract, C_{GLTHI}. Finally, Column (5) shows the percentage welfare difference between the GLTHI and the optimal contract; $\frac{C_{GHHW}-C_{GLTHI}}{C_{GHHW}}$.

Several findings emerge from Table 4. First, comparing Column (1) and Column (2), we find that a series of short-term contracts can produce large welfare losses compared to the first-best. For all initial health categories at age 25 and for both lifecycle income profiles, the CEs are negative. This highlights the significant negative welfare consequences of one-sided commitment, i.e., the inability of policyholders to commit to long-term contracts, together with the inability of consumers to borrow.[32] It is important to note, however, that the performance of short-term contracts is sensitive to catastrophic risk. Under the baseline scenario, individuals would pay \$54,930 whenever they reach category 7 under short-term contracts. This event, even if unlikely in a given year, has highly detrimental effects on welfare. We come back to this issue in Section 6.4.1.

Long-term contracts can produce substantial welfare gains compared to short-term contracts. Consider Panel (a) for the case where $\lambda_{25} = 1$ at inception. Column (3) shows that, under the optimal contract, the CE is \$22,448 for *Ed 10* and \$27,726 for *Ed 13*. Column (4) shows that, under the GLTHI contract, the CE is \$21,536 for *Ed 10* and \$26,024 for *Ed 13*.

Column (5) shows that the CE under the GLTHI contract is 4.2 (6.1, respectively) percent below than under the optimal contract for *Ed10* (*Ed13*, respectively) starting in the healthiest risk category ($_{25} = 1$). Comparing Columns (5) across panels (a) to (g), the welfare difference between the optimal and the GLTHI contract shrinks as health at inception worsens. It becomes almost negligible for individuals starting in health category 5 or worse. Welfare under both long-term contracts (GLTHI and GHHW) is highly driven by the large degree of frontloading when health status at inception deteriorates (see Table 3). In this case, both long-term contracts are ineffective in dealing with reclassification risk.

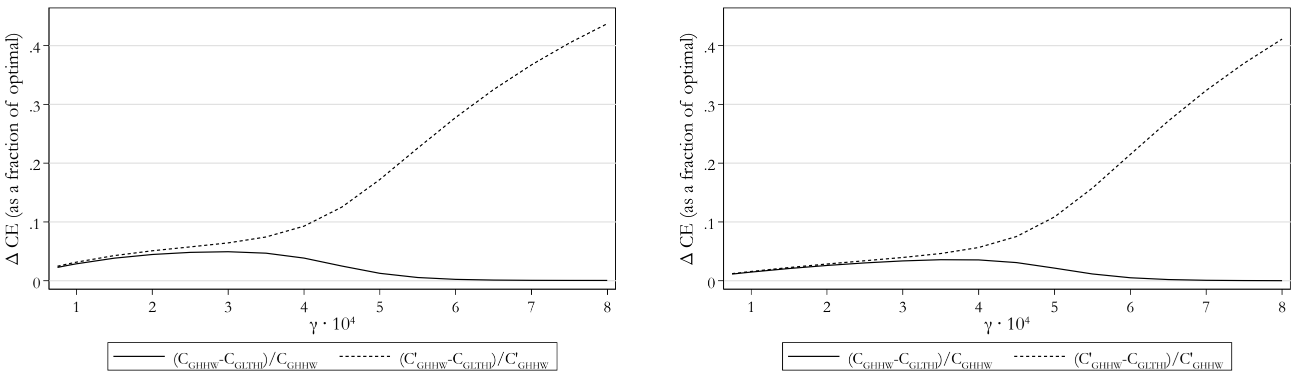
Finally, Panel (h) provides our baseline results using initial health distributions that correspond to the observed distributions at age 25 in our sample. As most individuals start healthy (99.35% have $\lambda_{25} \in \{1, 2\}$), both long-term contracts produce substantial gains relative to short-term contracts and recoup a large fraction of the welfare loss from short-term contracting. This highlights the large benefits of long-term contracts in providing insurance against reclassification risk. The welfare difference between the GLTHI and the optimal contract is close to four percent.

[32]As is well known, if consumers could borrow, they can "manufacture" commitment power by posting a "bond" with the insurer that equates the discounted sum of expected medical claims and restore first-best outcomes (see, e.g., Cochrane (1995) and Hendel and Lizzeri (2003)).

6.4.1 Robustness: Preferences, Initial Health Status, Savings, and Income Profiles

We investigate the robustness of our previous result in various dimensions. First, to different levels of risk aversion. Second, we use a CRRA utility function instead of CARA. Third, we use Epstein and Zin (1989)'s recursive preferences where risk aversion and intertemporal elasticity of substitution are separately parameterized. Fourth, we use claims data from one of the biggest German public insurers to replicate our main findings for an initial health status representative of the population of Germany, and we further investigate the sensitivity of the results to the initial distribution of health status. Fifth, we allow for savings. Finally, we use U.S. income profiles. We provide details of each robustness analysis below.

Risk Aversion Our baseline results assume risk aversion $\gamma = 4 \times 10^{-4}$. The solid lines in Figure 7 show the percentage point difference in CEs between each long-term contract as we vary the value for γ over the interval $\gamma \in [7.5 \times 10^{-5}, 8 \times 10^{-4}]$. Panel (a) shows the results for the flatter income profile (Ed 10), while Panel (b) for the steeper income profile (Ed 13).



(a) *Ed 13* (b) *Ed 10*

Figure 7: Welfare Difference (GLTHI vs GHHW) by Risk Aversion

Source: The figure shows the differences in Certainty Equivalents (CE) between GLTHI and the GHHW contract as a fraction of welfare under the GHHW contract, for different levels of risk aversion (γ). Panel (a) shows results for the steeper income profile (Ed 13), Panel (b) shows results for the flatter income profile (Ed 10). Source: Own calculations based on German Claims Panel Data and SOEP (2018).

We find that the percentage difference in welfare between the GHHW and GLTHI contracts is robust to the degree of risk aversion. The solid curves in the two panels show that the maximal welfare difference between the two contracts across all values of γ is 4.4 percent for Ed 10 and five percent for Ed 13, when $\gamma = 3.7 \times 10^{-4}$ or $\gamma = 3 \times 10^{-4}$, respectively. As expected, the welfare gap is small when γ is small, as reclassification risk and lack of intertemporal consumption smoothing are less relevant with less curvature in the utility function. However, the welfare gap is also small when γ is high.

To understand the mechanisms behind this pattern, we compute welfare under the lifecycle path of *expected* consumption under the GHHW contract—which we denote by C'_{GHHW}—and compare it to welfare under the GLTHI contract. In other words, we eliminate the welfare differences due to differences in reclassification risk, and quantify solely the welfare differences arising from better intertemporal consumption smoothing in the optimal contract. The dashed curves in Figure 7 plot $\frac{C'_{GHHW}-C_{GLTHI}}{C'_{GHHW}}$.[33]

For relatively low levels of γ, C'_{GHHW} is very close to C_{GHHW}, implying that the difference in CE between GLTHI and GHHW are almost exclusively determined by their differences in intertemporal consumption smoothing.[34] However, the role of reclassification risk in explaining the differences across contracts is more dominant at high values of γ. For instance, for $\gamma = 0.0004$ and for Ed 13, the average lifecycle consumption path under the optimal contracts produces welfare (C'_{GHHW}) that is 9.3 percent higher than welfare under GLTHI (compared to 3.9 percent higher when considering consumption volatility as in Panel (h) of Table 4). Overall, when γ is large, the welfare benefits of less reclassification risk under GLTHI largely compensate the welfare losses due to worse intertemporal consumption smoothing.

In summary, varying the level of risk aversion affects the performance of the GLTHI relative to the optimal contract via two underlying channels. The first is due to differences in intertemporal consumption smoothing, where GLTHI clearly falls short. The larger γ, the bigger this effect. The second is due to differences in reclassification risk, where GLTHI outperforms the optimal contract. Again, the larger γ, the bigger this effect, and the closer is welfare under GLTHI to welfare under the optimal contract.

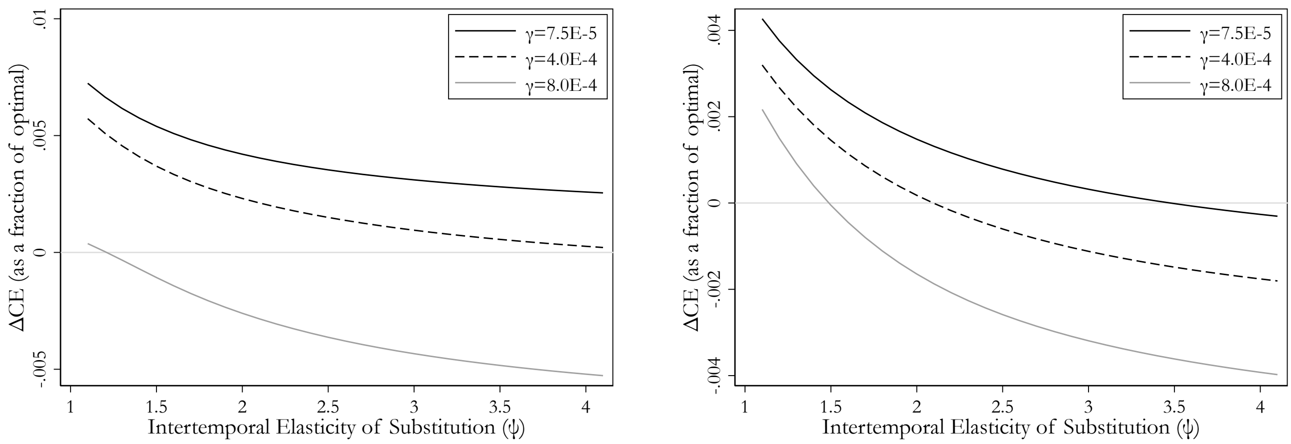
CRRA Preferences. Our main analysis assumes CARA preferences. We investigate the robustness of the findings to using CRRA preferences, that is $u(c) = \frac{u^{1-\sigma}}{1-\sigma}$. Previous research has suggested that CRRA might represent choices well (see e.g. Chiappori and Paiella 2011). Following Dohmen et al. (2011), we use a coefficient of risk aversion of $\sigma = 4$. We find welfare differences of 2.4% for Ed 10 and 3.8% for Ed 13.[35]

[33]In principle, one could also eliminate reclassification risk from the consumption path under GLTHI. However, the reclassification risk component of GLTHI is negligible.

[34]Note that even at these relatively low levels of γ, the lack of intertemporal consumption smoothing decreases welfare substantially, relative to risk neutrality. However, welfare decreases with γ by a similar amount when we eliminate reclassification risk. For instance, for Ed 13, CE_{GHHW} and $CE_{GHHW'}$ decrease by 14 percent and 13 percent compared to risk neutrality when $\gamma = 3 \times 10^{-4}$.

[35]To avoid negative consumption, we impose a consumption floor of \$10,000. This is a binding constraint in approximately 10 out of 10,000 simulated consumption levels.

Epstein-Zin Recursive Preferences. Our main specification assumes that a single parameter governs both risk aversion and the intertemporal elasticity of substitution. Here we break the parametric link between risk aversion, γ, and the intertemporal elasticity of substitution, ψ. As in Epstein and Zin (1989), preferences are defined recursively as $V_t = F(c_t, R_t(V_{t+1}))$ with $R_t(V_{t+1}) = G^{-1}(\mathbb{E}_t G(V_{t+1}))$, and we consider the CES aggregator $F(c,z) = \left((1-\delta)c^{1-1/\psi} + \delta z^{1-1/\psi}\right)^{\frac{1}{1-1/\psi}}$. We embed the same CARA specification used in our main analysis into the EZ preferences by assuming $G(c) = u(c) = \frac{1}{\gamma}e^{-\gamma c}$. Appendix A10.1 provides the details of the derivations.



(a) *Ed 13* (b) *Ed 10*

Figure 8: Welfare Differences (GHLTI vs. GHHW) with EZ preferences

Notes: The figure shows the differences in Certainty Equivalents (CE) between GLTHI and the GHHW contract as a fraction of welfare under the GHHW contract with recursive (Epstein-Zin) preferences, for different levels of intertemporal elasticity of substitution ψ, and for different levels of risk aversion (γ). Panel (a) shows results for the steeper income profile (Ed 13), Panel (b) shows results for the flatter income profile (Ed 10). Source: Own calculations based on German Claims Panel Data and SOEP (2018).

Varying γ and ψ, Figure 8 shows differences in CEs between the GLTHI and the optimal GHHW contract. Each panel corresponds to a different income profile. We show results for both extremes of the risk aversion interval in Figure 7 ($\gamma = 7.5 \times 10^{-5}$ and $\gamma = 8.0 \times 10^{-4}$), as well as our baseline risk aversion parameter ($\gamma = 4 \times 10^{-4}$). As expected, the welfare differences are larger with low levels of risk aversion and low intertemporal elasticity of substitution. The maximum welfare difference is 0.7%, which occurs for the steeper income profile. We also find that the GLTHI can even outperform the optimal contract when the intertemporal elasticity of substitution is relatively high.[36]

Initial Health Status Representative of Germany. Our results so far use the estimates of health status for the population of GLTHI policyholders. However, several institutional features imply that this population is not representative of the German population, and likely healthier (see Section 2).

[36] This can occur because the GHHW contract in Ghili et al. (2022) is not necessarily the optimal contract under recursive preferences—recall that Ghili et al. (2022)'s theoretical characterization requires that preferences are *time separable*, which Epstein and Zin (1989)'s recursive preferences do not satisfy.

This section studies the robustness of our results to using a health risk profile that is more representative of the German population as a whole. To do so, we use claims data from one of the biggest public insurers, with more than 5 million enrollees.

Overall, the claims data confirm that the privately insured are healthier. Figure A14 compares ACG risk score distributions across the publicly insured and the publicly insured; Table A9 compares average raw scores, and Table A10 shows the distribution by age over risk classes within the sample of publicly insured (all exhibits are in the Appendix).[37]

Using the risk scores and estimated initial probabilities at age 25 for the publicly insured, we find a welfare difference (GLTHI vs. GHHW) of 0.9 and 1.5 percent for Ed 10 and Ed 13, respectively. Appendix A10.2 provides additional details. [38]

Different Starting States. We further study the robustness of the results to the initial health status by considering a large number of draws of distributions over the initial health categories for the privately insured. Specifically, we sampled 20 million probability simplices $\tilde{\Delta}_r \in \Delta^7$ from a Dirichlet distribution with concentration parameters equal to the baseline probabilities for the privately insured.

We find that the welfare loss of the GLTHI relative to the optimal contract is bounded at around six percent for the highly educated *Ed13* and at around four percent for the intermediate educated *Ed10*. This exercise also confirms that the welfare differences between the GLTHI and the optimal contract is smaller when the population is less healthy at inception, which alleviates the concern that our findings are driven by a relatively healthy subsample of the overall population. Appendix A10.3 provides the details.

Savings. Our main welfare calculations assume that individuals cannot save. Savings could improve welfare under the GLTHI contract, particularly in settings with hump-shaped lifecycle income profiles like the ones we find in the SOEP. Here, we allow for savings, by solving the dynamic programming problem of optimal savings with mortality risk as in Yaari (1965). The details are in Appendix A10.4.

We find that the relative gap between GLTHI and the optimal contract is almost unaffected by savings. Intuitively, the GLTHI contract already forces more savings than desired through highly

[37] Furthermore, using the representative SOEP, Table A8 (Appendix) shows that privately insured are less likely to smoke, have lower BMIs, and use fewer health care services.

[38] Another question is whether risk tolerance differs between publicly and privately insured. At least in West Germany, there is evidence that civil servants are more risk averse than the rest of the population (Fuchs-Schündeln and Schündeln, 2005). Table A8 confirms this. Using the SOEP, Figure A13 plots very similar risk tolerance distributions for both populations. If anything, privately insured are a bit less risk averse, mostly stemming from more mass between 6 and 8 on the 1-10 Likert scale.

front-loaded premiums. Moreover, under the optimal contract, individuals have no incentives to save additionally as shown in Ghili et al. (2022). Thus, savings do not affect welfare under the optimal contract.

Income Profiles. Finally, to test the robustness with respect to the lifecycle income profile, we now employ the representative Panel Study of Income Dynamics (PSID) for the United States.[39] To extract lifecycle income profiles, we use the exact same income concept as in our main analysis for German income profiles. Moreover, we use the same estimation process. That is, we exclude respondents under 25, focus on the years 1984 to 2015, and estimate Equation (6) for high-school and college educated. Figure A16 (Appendix) shows the estimated lifecycle income profiles.

In the U.S., the increase in post-tax equivalized income between ages 25 and 60 is very close to the pattern in Germany. However, the decrease in income after age 60 is much steeper in the U.S., for both educational groups. This effect amplifies the beneficial effects of optimal contracts in allowing for consumption smoothing at old ages, as shown in Figure 5. Overall, the GLTHI contract would achieve welfare that would fall 5.8 and 3.5 percent short of the optimal dynamic long-term contract for Americans with high school and college degrees, respectively. To illustrate the role of the hump-shaped income, we calculate the welfare gap between the GLTHI and GHHW contracts under the PSID income for long-term contracts that would expire at 65 years old.[40] In such case, the welfare gap between the GLTHI and GHHW contracts is 2.3 percent and 2.0 percent for high school and college income profiles, respectively. Also, allowing for savings in life-long contracts, the welfare gap is 3.7 percent and 3.0 percent for high school and college income profiles, respectively.

6.5 Benchmarking the Loss from Using Flat-Income Optimal Contracts

In our setting, the flat-income optimal GLTHI contracts provide welfare that is around four percent lower than with the optimal contracts. Here we put this welfare difference in perspective, and provide additional lessons regarding when and why the flat-income optimal GHTHI contract can approximate the welfare gains from the optimal contracts.

The benefits of replacing short term contacts with each long-term contract provide a natural benchmark to compare the welfare differences across both long-term contracts. Panel (a) of Table 5 expands our main results by providing additional metrics to assess the welfare gains of long-term

[39] The PSID is the oldest and longest-running panel survey in the world. From 1968 to 1996, it surveyed the U.S. families annually, and from 1997, biannually (Panel Study of Income Dynamics, 2018). We use the Cross National Equivalence Files (CNEF), which harmonizes survey measures across years (Frick et al., 2007).

[40] This exercise is also interesting from a policy perspective if long-term contracts were considered for the US working population while keeping the Medicare program intact.

over short-term contracts. Columns (1) to (4) replicate the main results for CE under each contract. Column (5) and (6) show what fraction of the welfare difference between the first best and the short-term contract is recouped by the optimal contract and the GLTHI, respectively; $\frac{C_{GHHW}-C_{ST}}{C^{*}-C_{ST}}$ and $\frac{C_{GLTHI}-C_{ST}}{C^{*}-C_{ST}}$. Column (7) shows the fraction of the gains from replacing short-term contracts with the optimal (GHHW) contracts that is achieved by replacing short term contract with the GLTHI contract; $\frac{C_{GLTHI}-C_{ST}}{C_{GHHW}-C_{ST}}$. Finally, Column (8) shows again the percentage welfare difference between the GLTHI and the optimal contract.

Under our baseline results, the optimal contracts are very effective. They close either 96.9 percent or 77.8 percent of the welfare gap between short-term contracts and the first best for each income profile, respectively (Column 5). This highlights the large benefits of optimal long-term contracts in providing insurance against reclassification risk. Similarly, the GLTHI contract is able to close either 94.5 percent or 75.1 percent of that gap (Column 6). As a consequence, for both income profiles, the GLTHI contract achieves close to 97 percent of the welfare gains from optimal long-term contracting (Column 7). Even though the GLTHI contract is only optimal when income is constant over the life-cycle, its ability to insure *large* reclassification risk, and its particular balance between front-loading and reclassification risk insurance still achieves much of the benefits from optimal long-term contracting.

Table 5: Benchmarking Welfare under GLTHI

	(1) C^{*}	(2) C_{ST}	(3) C_{GHHW}	(4) C_{GLTHI}	(5) $\frac{C_{GHHW}-C_{ST}}{C^{*}-C_{ST}}$	(6) $\frac{C_{GLTHI}-C_{ST}}{C^{*}-C_{ST}}$	(7) $\frac{C_{GLTHI}-C_{ST}}{C_{GHHW}-C_{ST}}$	(8) $\frac{C_{GHHW}-C_{GLTHI}}{C_{GHHW}}$
	Panel (a): Baseline Results							
Ed 10	22,980	-10,119	21,945	21,168	96.9%	94.5%	97.6%	3.5%
Ed 13	34,159	-2,223	26,093	25,088	77.8%	75.1%	96.5%	3.9%
	Panel (b): Without Catastrophic Losses							
Ed 10	22,979	14,504	21,930	21,164	87.6%	78.6%	89.7%	3.5%
Ed 13	34,159	22,123	26,064	25,079	32.7%	24.6%	75.0%	3.8%

Notes: The table shows welfare measured by the Consumption Certainty Equivalents in 2016 USD dollars, per capita, per year, separately for two income profiles (see Figure 3). Panels (a) uses the baseline contracts as in Table 4. Panel (b) assumes that catastrophic risk is insured by an alternative contract, as explained in the text. Columns (1) to (4) show welfare according to the (1) first-best (C^{*}), (2) a series of short-term contracts (C_{ST}), (3) the optimal contract (C_{GHHW}), and (4) the GLTHI (C_{GLTHI}). Column (5) and (6) show how much of the welfare gap between (2) and (1) is closed by the optimal contract and the GLTHI, respectively. Column (7) shows the fraction of the potential benefits from long-term contracting achieved by GLTHI. Column (8) shows the percentage of welfare loss under GLTHI relative to the optimal contract. Source: Own calculations based on German Claims Panel Data, SOEP (2018).

The Role of Catastrophic Risk. As previously noted, welfare under short-term contracts is very low in our baseline scenario. Our risk classification method accommodates the heavily-skewed distribution of expenditures with a high-cost but low-probability risk category 7 (Section 5.1). Even though we observe large health expenditures in our data, the assumption that individuals would be subject to such large losses in the counterfactual scenario with short-term contracts may be unrealistic, for instance, due to public safety-net programs, or other forms of protection against catastrophic risks. Relatedly, one may wonder whether *any* mechanism that provides insurance against this catastrophic risk would perform close to the optimal dynamic contract.

We now evaluate the welfare effects under different contracts when catastrophic losses are bound. Specifically, we recompute welfare under each contract by capping the expenditures of category 7 at those of category 6. To finance that cap, individuals starting at age 25 in $\lambda_{25} = k_0$ pay a yearly fee $\tilde{P}(k_0)$ such that the insurance company breaks-even ex ante.[41] In particular, premiums under this modified short-term contract are now $P_t(\lambda_t|\lambda_{25} = k_0) = \tilde{P}(k_0) + \mathbb{E}(m_t|\min\{\lambda_t, 6\}|\lambda_{25} = k_0)$. The long-term contracts are also computed as before, but with the capped health expenditures and the additional yearly fee.[42] Panel (b) of Table 5 provides the results under this alternative set of contracts.

We find that the gap between the two long-term contracts (GLTHI and GHHW) is robust to catastrophic expenditures (Column 8 of Table 5). It remains below four percent after we eliminate the risk of category-7 expenditures. In fact, welfare under both long-term contracts changes little under this alternative scenario. This is because long-term contracts have the virtue of preventing individuals from paying higher premiums when they transition to high-cost categories. In particular, health expenditures of category 7 only affect welfare by slightly affecting the contract terms (as state 7 only arises with very low probability in the data and the contract terms are set by the risk-neutral insurer). This result highlights that long-term contracts can be very effective in dealing with the risk of future catastrophic losses.[43]

As expected, welfare under short-term contracts increases substantially when we limit the size of the catastrophic risk. As a consequence, bounding catastrophic losses generally results in smaller gains from the optimal GHHW contracts. In particular, Column (5) shows that for the steeper income profile (Ed 13), optimal long-term contracts close 32.7 percent of the gap between the first best and short-term contracts, as opposed to 77.8 percent in our baseline scenario.

[41] That is, $\tilde{P}(k_0) = \frac{\mathbb{E}\left(\sum_{t=25}^{T} S_t \delta^{t-25} \mathbb{1}[\lambda_t=7](\mathbb{E}(m_t|\lambda_t=7) - \mathbb{E}(m_t|\lambda_t=6))|\lambda_{25}=k_0\right)}{\mathbb{E}\left(\sum_{t=25}^{T} S_t \delta^{t-25}|\lambda_{25}=k_0\right)}$ where $\mathbb{1}[]$ is the indicator function.

[42] We note that these alternative contracts are not "incentive-compatible," as individuals who become healthier than expected over time could switch to an alternative contract with a lower fee. The proposed contracts assume commitment in the payment of $P_t(\lambda_t|\lambda_{25} = k_0)$.

[43] It is also interesting to note that welfare under both long-term contracts decreases (slightly) in this alternative scenario, even if we assume that individuals can commit to paying the premium $P_0(\lambda_{25})$ over time, regardless of the evolution of their health status.

Naturally, capping catastrophic losses increases the welfare gap from using the GLTHI instead of the optimal GHHW when measured as a fraction of the gains from optimal long-term relative to short-term contracting. The reason is simple: The welfare gap between the GLTHI and GHHW long-term contracts stays almost unaltered in *absolute* terms, but the gains from optimal long-term contracts relative to short-term contracting shrinks. In our baseline scenario with large catastrophic losses, GLTHI recoups 97.6 percent (for Ed 10) and 96.5 percent (for Ed 13) of the gains from replacing short term contracts with the optimal long-term contracts (Column (7), Panel (a), Table 5). However, after bounding category-7 expenses, the gains from optimal long-term contracting are more limited, and GLTHI recoups 89.7 percent (for Ed 10) and 75.0 percent (for Ed 13) of those (smaller) gains (Column (7), Table 5).

Both long-term contracts still bring sizable gains beyond those that come from capping catastrophic risk, particularly for flatter income profiles (CE increases from 14,504 to 21,164 under the GLTHI contract and to 21,930 under the GHWW contract). It is also worthwhile noting that, in our empirical setting, both long-term contracts improve welfare substantially relative to the "guaranteed-renewable" contracts of Pauly et al. (1995), which fully eliminate reclassification risk at the expense of even higher frontloading.[44] These results highlight the virtue of optimal long-term contracting, in that their welfare gains do depend on the intricate trade-offs between intertemporal consumption smoothing and reclassification risk insurance. The particular way in which GLTHI balances this tradeoff, and its shared commonalities with the optimal contract, are important to explain its performance.

The results from eliminating the catastrophic risk are tightly connected to the fact that the German contract provides more reclassification risk insurance at the expense of more frontloading than the optimal dynamic contract. In our setting, much of the welfare loss from short-term contracting would be due to reclassification risk.[45] Under such situations, both the GLTHI and the optimal GHHW contracts are very effective as they are designed to mitigate reclassification risk using frontloaded premiums. By virtue of practically eliminating reclassification risk, the GLTHI contract achieves much of the benefits of long-term contracting, even though it introduces more frontloading than optimal. In contrast, when the size of the catastrophic risk is bounded, a smaller fraction of the losses from short-term contracting is due to reclassification risk (and a larger fraction is due to the lack of

[44] Results are available from the authors upon request.

[45] To make this point more precisely, we can quantify the relative importance of reclassification risk under our benchmark scenario as well as under the alternative assumptions regarding catastrophic losses by computing the certainty consumption equivalent under a benchmark of no-borrowing and no-savings (C_{NBNS}) and calculating $\frac{C_{NBNS}-C_{ST}}{C^{*}-C_{ST}}$. This metric quantifies what fraction of the overall welfare losses from short-term contracts $(C^{*} - C_{ST})$ are due to reclassification risk $(C_{NBNS} - C_{ST})$. The remainder is due to the lack of intertemporal consumption smoothing. We find that in our baseline scenario, for Ed 10 (Ed 13), this metric is equal to 97.3 percent (82.7 percent), whereas after bounding the category-7 catastrophic losses it is 89.8 percent (47.4 percent).

intertemporal consumption smoothing).

The difference across income groups in the relative performance of each contract also relates directly to the source of welfare losses under short-term contracting. For relatively flat income profiles (Ed 10), most of the welfare loss from short-term contracting is due to reclassification risk, even when the size of the catastrophic risk is bounded. As a consequence, the GLTHI contract achieves a large fraction of the potential gains from optimal long-term contracting. For steeper income profiles, however, a larger fraction of the welfare gap comes from poor intertemporal consumption smoothing, especially when catastrophic risk is bounded. In these cases, long-term contracts are in general less effective; and the GLTHI contract achieves a smaller fraction of the potential gains from optimal long-term relative to short-term contracting.

6.6 Summary

In our setting, we find that flat-income optimal design achieves welfare that is close to 96 percent of the welfare under the optimal contract. This result is robust to several assumptions.

When reclassification risk is high, long-term contracts are very effective and provide large welfare gains relative to short term contracts. In such cases, the welfare gains of replacing short term contracts with the flat-income optimal design represent most of the gains from optimal long-term contracting. However, when catastrophic losses are more limited, the gains from optimal long-term contracting are also more limited, and it is also the case that the flat-income optimal recoups a smaller share of the gains. This pattern is tightly linked to the fact that the flat-income optimal contract provides more reclassification risk insurance at the expense of worse intertemporal consumption smoothing.

7 Conclusion

Regulated private markets for comprehensive health insurance have an important role in several OECD countries—including the U.S., Chile, Switzerland, and Germany. A central goal of premium rating regulation in such markets is to prevent reclassification risk, which is often achieved through community-rated pricing. For example, Switzerland has traditionally organized its private health insurance market with short-term annual contracts with tight community-rating regulations to provide reclassification risk insurance for all citizens. Several large markets and programs in the U.S. also feature community rating, most notably the individual market after the enactment of the Affordable Care Act (ACA). However, community rating regulations, unless combined with individual mandates, tend to exacerbate adverse selection (Geruso and Layton, 2017).

A fundamental, but empirically understudied alternative to regulated short-term contracts are individual *long-term* contracts. They provide policyholders with reclassification risk insurance without necessarily triggering adverse selection. This paper bridges the theoretical literature on long-term contracts with its most important real-world application—the German individual private long-term health insurance market (GLTHI).

We show that contracts in Germany's private health insurance share several features with the optimal dynamic contract as recently derived by Ghili et al. (2022). Moreover, we highlight that the two contracts coincide for individuals with a flat lifecycle income profile. More generally, the GLTHI contract almost fully eliminates reclassification risk over a policyholder's lifecycle. However, the elimination of reclassification risk comes at the expense of higher premium front-loading which, compared to the optimal contract, results in less intertemporal consumption smoothing.

We assess welfare under the GLTHI contracts and compare it to the optimal contracts by leveraging unique claims panel data from GLTHI policyholders along with household panel data. Key to this analysis, we propose a novel risk classification method based on actuarial science principles. We find that welfare under the GLTHI design—as measured by certainty-equivalent yearly consumption—is 96 percent of what the income-dependent optimal contract achieves. This result is robust to the degree of risk aversion, to using non-time-separable recursive preferences à la Epstein and Zin (1989), to using the initial health status of the publicly insured in Germany, to using U.S. lifecycle income profiles, and to the extent of catastrophic risk. Overall, the welfare loss due to less consumption smoothing from ignoring income dynamics is partly compensated by less reclassification risk.

We also provide further insights regarding the workings of the flat-income-optimal GLTHI contract relative to the income-dependent optimal dynamic contracts. In our empirical setting, a large share of the welfare loss under the counterfactual scenario of short-term contracts would be due to reclassification risk. The flat-income-optimal and optimal contracts are very effective at improving upon short-term contracts: they mitigate the large reclassification risk with premium frontloading in a way that results in large welfare gains over short-term contracts. However, in scenarios with more limited catastrophic risk, e.g. if long-term contracts were to coexist with government programs that already insure catastrophic expenses, the additional gains from optimal long-term contracting would be smaller, and the GLTHI design would recoup a smaller share of those incremental gains.

A practical advantage of the GLTHI contract is its simplicity in that, after an initial risk rating, they are guaranteed renewable with constant and income-independent premiums as long as the policyholder decides to renew. This simplicity may explain its proliferation and stable existence in Germany for decades. We see the good performance of long-term contracts under this particular design as a virtue of the theory of optimal long-term contracts, which strengthens their case as an appealing policy option. In the U.S., so far, the debate has largely focused on either incremental adjustments to the community-rated short-term contracts under the ACA, or the transition to a "single-payer for all" system.

Finally, we would like to acknowledge two important and general caveats of long-term contracts. First, our results show that neither the GLTHI nor the optimal dynamic contract are affordable for those who are sick in young ages. From a policy perspective, societies implementing long-term contracts must offer a solution for those individuals. Second, our discussion abstracts from a couple of key features that may have implications for welfare under the long-term contracts. First, our model assumes time-consistent individuals. From the perspective of a present-biased consumer, front-loading may render long-term contracts undesirable, particularly when front-loading is high.[46] In addition, our model abstracts from moral hazard. In the presence of moral hazard, using long-term contracts to protect against reclassification risks could also induce inefficiencies in health spending and health investment, similar to what is studied in Cole et al. (2019) for the case of community rating. Quantifying the role of moral hazard in long-term contracts is an important avenue for future research.

References

Aizawa, N. and H. Fang (2020). Equilibrium labor market search and health insurance reform. *Journal of Political Economy 128*(11), 4258–4336.

Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics 84*(3), 488–500.

Association of German Private Healthcare Insurers (2020). *Zahlenbericht der Privaten Krankenversicherung 2019*. Verband der Privaten Krankenversicherung. https://www.pkv.de/fileadmin/user_upload/PKV/c_Verband/PDF/2020-12_PKV-Zahlenbericht_2019.pdf, retrieved on January 18, 2021.

Association of German Private Healthcare Insurers (2022). *PKV Zahlenportal*. Verband der Privaten Krankenversicherung. https://www.pkv-zahlenportal.de/werte/2010/2020/12, retrieved February 20, 2022.

Atal, J., H. Fang, M. Karlsson, and N. R. Ziebarth (2019). Exit, voice or loyalty? An investigation into mandated portability of front-loaded private health plans. *Journal of Risk and Insurance 86*(3), 697–727.

Atal, J. P. (2019). Lock-in in dynamic health insurance contracts: Evidence from Chile. PIER Working Paper Archive 19-020, Penn Institute for Economic Research, Department of Economics, University of Pennsylvania.

[46] Still, Gottlieb and Zhang (2019) show that with a sufficiently large number of periods, the inefficiencies arising from time inconsistency vanish. With the long-term contract that emerges in the equilibrium with time-inconsistent agents, time-inconsistent agents may achieve the same level of welfare as time-consistent agents.

Athey, S. and G. W. Imbens (2019). Machine learning methods that economists should know about. *Annual Review of Economics 11*(1), 685–725.

Bahnsen, L. and F. Wild (2021). Entwicklung der prämien- und beitragsanpassungen in pkv und gkv 2012-2022. Wip-kurzanalyse, Wissenschaftliches Institut der PKV. http://www.wip-pkv.de/forschungsbereiche/detail/entwicklung-der-praemien-und-beitragseinnahmen-in-pkv-und-gkv-2012-2022.html, retrieved February 10, 2022.

Baumann, F., V. Meier, and M. Werding (2008). Transferable ageing provisions in individual health insurance contracts. *German Economic Review 9*(3), 287–311.

Becker, G. S. and B. R. Chiswick (1966). Education and the distribution of earnings. *The American Economic Review 56*(1/2), 358–369.

Bhuller, M., M. Mogstad, and K. G. Salvanes (2017). Life-cycle earnings, education premiums, and internal rates of return. *Journal of Labor Economics 35*(4), 993–1030.

Börsch-Supan, A. and H. Jürges (2012). Disability, pension reform, and early retirement in Germany. In *Social Security Programs and Retirement around the World: Historical Trends in Mortality and Health, Employment, and Disability Insurance Participation*, NBER Chapters, pp. 277–300. National Bureau of Economic Research.

Bundesministerium für Gesundheit (2022). Gesundheitsberichterstattung des Bundes. http://www.gbe-bund.de/, retrieved on February 6, 2022.

Bünnings, C., H. Schmitz, H. Tauchmann, and N. R. Ziebarth (2019). How health plan enrollees value prices relative to supplemental benefits and service quality. *Journal of Risk and Insurance 86*(2), 415–449.

Büser, W. (2012). Private Krankenversicherung: Der Weg zurück in die Kasse. *Deutsches Ärzteblatt 109*(4), A–162–A–162.

Cecu (2018). *Rückkehr aus der privaten in die gesetzliche Krankenversicherung Überblick zum PKV-GKV-Wechsel*. Portal für Finanzen und Versicherungen. https://www.cecu.de/private-krankenversicherung-rueckkehr-gesetzliche.html, last retrieved on October 28, 2018.

Chiappori, P.-A. and M. Paiella (2011). Relative risk aversion is constant: Evidence from panel data. *Journal of the European Economic Association 9*(6), 1021–1052.

Christiansen, M., M. Eling, J.-P. Schmidt, and L. Zirkelbach (2016). Who is changing health insurance coverage? Empirical evidence on policyholder dynamics. *Journal of Risk and Insurance 83*(2), 269–300.

Claxton, G., L. Levitt, and K. Pollitz (2017). *Pre-ACA Market Practices Provide Lessons for ACA Replacement Approaches*. Kaiser Family Foundation. https://www.kff.org/health-costs/issue-brief/pre-aca-market-practices-provide-lessons-for-aca-replacement-approaches/, retrieved on September 14, 2018.

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

Cole, H. L., S. Kim, and D. Krueger (2019). Analysing the effects of insuring health risks: On the trade-off between short-run insurance benefits versus long-run incentive costs. *Review of Economic Studies 86*(3), 1123–1169.

Deutsche Aktuarvereinigung (DAV) (2019). *Der aktuarielle Unternehmenszins in der privaten Krankenversicherung (AUZ): Richtlinie*. https://aktuar.de/unsere-themen/fachgrundsaetze-oeffentlich/2019-10-09_DAV-Richtlinie_AUZ.pdf, retrieved on February 6, 2010.

Diamond, P. (1992). Organizing the health insurance market. *Econometrica 60*(6), 1233–54.

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. G. Wagner (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the european economic association 9*(3), 522–550.

Dustmann, C., P. A. Puhani, and U. Schönberg (2017). The long-term effects of early track choice. *The Economic Journal 127*(603), 1348–1380.

Eekhoff, J., M. Jankowski, and A. Zimmermann (2006). Risk-adjustment in long-term health insurance contracts in Germany. *The Geneva Papers on Risk and Insurance-Issues and Practice 31*(4), 692–704.

Einav, L., A. Finkelstein, S. P. Ryan, P. Schrimpf, and M. R. Cullen (2013). Selection on moral hazard in health insurance. *American Economic Review 103*(1), 178–219.

Epstein, L. G. and S. E. Zin (1989). Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework. *Econometrica 57*(4), 937–969.

Fang, H. and E. Kung (2020). Life insurance and life settlements: The case for health-contingent cash surrender values. *Journal of Risk and Insurance 87*(1), 7–39.

Finger, R. J. (2001). Risk classification. In Casualty Actuarial Society (Ed.), *Foundations of Casualty Actuarial Science* (fourth ed.)., Chapter 6, pp. 287–341.

Finger, R. J. (2006). Risk classification. *Foundations of Casualty Actuarial Science*, 231–276.

Finkelstein, A., K. McGarry, and A. Sufi (2005). Dynamic inefficiencies in insurance markets: Evidence from long-term care insurance. *The American Economic Review 95*(2), 224–228.

Fleitas, S., G. Gowrisankaran, and A. L. Sasso (2018). Reclassification risk in the small group health insurance market. Technical report, National Bureau of Economic Research.

French, E. and E. Kelly (2016). Medical spending around the developed world. *Fiscal Studies 37*(3-4), 327–344.

Frick, J. R., S. P. Jenkins, D. R. Lillard, O. Lipps, and M. Wooden (2007). The Cross-National Equivalent File (CNEF) and its Member Country Household Panel Studies. *Journal of Applied Social Science Studies (Schmollers Jahrbuch: Zeitschrift für Wirtschafts- und Sozialwissenschaften) 127*(4), 626–654.

Fuchs-Schündeln, N. and M. Schñdeln (2005). Precautionary savings and self-selection: Evidence from the german reunification "experiment". *Quarterly Journal of Economics 120*(3), 1085–1120.

Geruso, M. and T. J. Layton (2017, November). Selection in health insurance markets and its policy remedies. *Journal of Economic Perspectives 31*(4), 23–50.

Geyer, J. and V. Steiner (2014). Future public pensions and changing employment patterns across birth cohorts. *Journal of Pension Economics and Finance 13*(02), 172–209.

Ghili, S., B. Handel, I. Hendel, and M. D. Whinston (2022). Optimal long-term health insurance contracts: Characterization, computation, and welfare effect. *The Review of Economic Studies*.

GKV-Gesundheitsreformgesetz (2000). Gesetz zur Reform der gesetzlichen Krankenversicherung ab dem Jahr 2000. `https://www.bgbl.de/xaver/bgbl/start.xav?start=%2F%2F*[%40attr_id%3D%27bgbl199s2626.pdf]`, last retrieved on February 10, 2020.

GKV-Spitzenverband (2022). Die Gesetzlichen Krankenkassen. `https://www.gkv-spitzenverband.de/krankenversicherung/kv_grundprinzipien/alle_gesetzlichen_krankenkassen/alle_gesetzlichen_krankenkassen.jsp`, retrieved on February 6, 2022.

Goebel, J., M. M. Grabka, S. Liebig, M. Kroh, D. Richter, C. Schröder, and J. Schupp (2019). The german socio-economic panel (soep). *Jahrbücher für Nationalökonomie und Statistik 239*, 345–360.

Gotthold, K. and B. Gräber (2015). So kommen Sie zurück in die gesetzliche Kasse. https://www.welt.de/finanzen/verbraucher/article138148142/So-kommen-Sie-zurueck-in-die-gesetzliche-Kasse.html, last retrieved on October 18, 2018.

Gottlieb, D. and X. Zhang (2019). Long-term contracting with time-inconsistent agents. mimeo. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3176354, retrieved February 15, 2020.

Handel, B., I. Hendel, and M. D. Whinston (2015). Equilibria in health exchanges: Adverse selection vs. reclassification risk. *Econometrica 83*(4), 1261–1313.

Harris, M. and B. Holmstrom (1982). A theory of wage dynamics. *The Review of Economic Studies 49*(3), 315–333.

Heckman, J. and B. Burton Singer (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica 52 (2)*, 271–320.

Hendel, I. and A. Lizzeri (2003). The role of commitment in dynamic contracts: Evidence from life insurance. *The Quarterly Journal of Economics 118*(1), 299–328.

Henke, K. D. (2007). Zur dualität von gkv und pkv / the future of private and public health insurance in germany. *Jahrbücher für Nationalökonomie und Statistik 227*(5-6), 502–528.

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

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

Hullegie, P. and T. J. Klein (2010). The effect of private health insurance on medical care utilization and self-assessed health in Germany. *Health Economics 19*(9), 1048–1062.

Innungskrankenkasse Berlin Brandenburg (2018). *IKK- oder privat versichert: Wichtige Infos zur Entscheidung*. https://www.ikkbb.de/fileadmin/user_upload/doc/Broschueren_fuer_Versicherte/IKKBBoderprivatversichertVers..pdf, last retrieved on October 30, 2018.

Kaiser Family Foundation (2019). *2019 Employer Health Benefits Survey*. https://www.kff.org/health-costs/report/2019-employer-health-benefits-survey/, retrieved on December 14, 2019.

Keane, M. P. and K. I. Wolpin (1997). The career decisions of young men. *Journal of Political Economy 105*(3), 473–522.

Kluth, S. and M. Gasche (2016). Ersatzraten in der Gesetzlichen Rentenversicherung / Replacement Rates in the German Statutory Pension System. *Jahrbücher für Nationalökonomie und Statistik (Journal of Economics and Statistics) 235*(6), 553–583.

Koijen, R. S. J., T. J. Philipson, and H. Uhlig (2016). Financial Health Economics. *Econometrica 84*, 195–242.

Krankenkassen-Zentrale (KKZ) (2020). *BAP 2020: Neue Private Krankenversicherung – Beitragserhöhung und andere Beitragsanpassungen*. https://www.krankenkassenzentrale.de/wiki/beitragsanpassung#, retrieved on February 6, 2010.

Kriegel, H.-P., E. Schubert, and A. Zimek (2017). The (black) art of runtime evaluation: Are we comparing algorithms or implementations? *Knowledge and Information Systems 52*(2), 341–378.

Krueger, D. and H. Uhlig (2006). Competitive risk sharing contracts with one-sided commitment. *Journal of Monetary Economics 53*(7), 1661–1691.

Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory 28*(2), 129–137.

Nell, M. and S. Rosenbrock (2007). Die Diskussion über die Portabilität von risikogerechten Transferbeträgen in der Privaten Krankenversicherung. *Zeitschrift für die gesamte Versicherungswissenschaft 96*(1), 39–51.

Nell, M. and S. Rosenbrock (2009). Ein Risikoausgleichsmodell für den Wettbewerb um Bestandskunden in der PKV. *Zeitschrift für die gesamte Versicherungswissenschaft 98*(4), 391.

Panel Study of Income Dynamics (2018). *Public use dataset*. University of Michigan, Ann Arbor, MI. Produced and distributed by the.

Panthöfer, S. (2016). Risk selection under public health insurance with opt-out. *Health Economics 25*(9), 1163–1181.

Patel, V. and M. V. Pauly (2002). Guaranteed renewability and the problem of risk variation in individual health insurance markets. *Health Affairs Suppl. Web Exclusives*, W280–W289.

Pauly, M., A. Percy, and B. Herring (1999). Individual versus job-based health insurance: Weighing the pros and cons. *Health Affairs 18*(6), 28–44.

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

Pauly, M. V. and R. D. Lieberthal (2008). How risky is individual health insurance? *Health Affairs 27*(3), w242–w249.

Polyakova, M. (2016). Risk selection and heterogeneous preferences in health insurance markets with a public option. *Journal of Health Economics 49*, 153 – 168.

Rosenbrock, S. (2010). *Einflussfaktoren für die Entscheidung bei Altersvorsorgeprodukten*.

Schencking, F. (1999). *Private Krankenversicherung im Unterschied zur Gesetzlichen Krankenversicherung*, pp. 18–32. Wiesbaden: Deutscher Universitätsverlag.

Schmeiser, H., T. Störmer, and J. Wagner (2014). Unisex insurance pricing: Consumers' perception and market implications. In *The Geneva Papers on Risk and Insurance—Issues and Practice*, Volume 39, pp. 102–138.

Schmidt, T. I. (2020). Die beihilfesysteme des bundes und der länder im vergleich. Rechtswissenschaftliches gutachten im auftrag des bundesministeriums fr̈ gesundheit, Bundesministerium für Gesundheit. `https://www.bundesgesundheitsministerium.de/service/publikationen/details/vergleich-der-beihilfesysteme-der-bundeslaender-und-des-bundes-aus-rechtlicher-sicht.html`, retrieved February 10, 2022.

Schmitz, H. and M. Westphal (2017). Informal care and long-term labor market outcomes. *Journal of Health Economics 56*(C), 1–18.

Schmitz, H. and N. R. Ziebarth (2017). Does framing prices affect the consumer price sensitivity of health plan choice? *The Journal of Human Resources 52*(1), 89–128.

SOEP (2018). *Data for years 1984-2016*. version 33, SOEPlong, 2018 doi:10.5684/soep.v33.

Statistisches Bundesamt (2020). Sozialleistungen - Angaben zur Krankenversicherung (Ergebnisse des Mikrozensus) 2019. Fachserie 13 reihe 1.1.

Werbeck, A., A. Wb̈ker, and N. R. Ziebarth (2021). Cream skimming by health care providers and inequality in health care access: Evidence from a randomized field experiment. *Journal of Economic Behavior & Organization 188*, 1325–1350.

Yaari, M. E. (1965). Uncertain lifetime, life insurance, and the theory of the consumer. *Review of Economic Studies 32*(2), 137–150.

Online Appendix

A1 Joining the PKV System and Enrolling in GLTHI

As mentioned in Section 2, the decision to join the PKV is essentially a lifetime decision. The basic social insurance principle is: "Once private, always private[ly insured]."

There are several reasons why people leave the GKV permanently to join the PKV system: First, PKV premiums are actuarially fair instead of income-dependent contribution rates. On the other hand, GKV offers free family coverage for non-working relatives. This implies that childless and healthy high income earners tend to be better off financially in the PKV.[A47]

Second, in the PKV system, people can choose between thousands of plans which they can customize according to individual preferences and needs. Applicants can choose their benefit level and cost-sharing amounts, within some lax regulatory limits.[A48] Most insurers operate nationwide, are open to all applicants, and sell policies in all states. Policies are also portable across state lines.

Third, the PKV relies on private contracts with one-sided commitment over the lifecycle. GKV pricing and benefits depend inherently on the political situation and funding. One could argue that the public system carries more uncertainty about future benefits and contribution rates. Below we discuss the GLTHI lifecycle premium calculation in theory and practice.

Fourth, the regulation makes the GLTHI particularly attractive for civil servants. The large majority of them enroll in it. The reason is that, since the 1920s, German law stipulates that the state has the obligation to care for their civil servants (*Fürsorgeplicht*), which is why the government self-insures (at least) 50 percent of their civil servants' health care costs directly (*Beihilfe*).[A49] However, in general, the government does not pay half the GKV contribution rates as private employers do (Schmidt, 2020). This implies that civil servants would have to pay the full GKV contribution rate, but only insure 50 percent of their health care costs on the PKV market which makes the latter much more attractive.[A50]

[A47]Obviously, the advantage of having income-independent premiums can backfire when income drops later in life. For financial hardship cases, the regulator introduced the Basic Plan, see footnote A56.

[A48]For example, effective 2009, the government mandated that deductible cannot exceed €5,000 for new policies (see *GKV-Wettbewerbsstärkungsgesetz*).

[A49] In practice, this means that civil servants, first have to pay providers out-of-pocket and then submit their claims to a state agency *and* the private insurer, both of which then reimburse half the costs. This principle that the insured must first pay providers and then submit claims (*Kostenerstattungsprinzip*) is a general GLTHI principle and downside for policyholders. Compared to GKV, where providers directly submit claims to sickness funds and cost-sharing is low, it implies more paperwork, uncertainty about claim denials, and requires more cash liquidity.

[A50] The exact share of *Beihilfe* depends on the state, see Schmidt (2020) for details. Several private insurers have specialized in providing customized coverage for the remaining uninsured share. One out of 16 states, Bremen, also generally offers civil servants the option to get half of their GKV contribution rate covered, which essentially implies giving them

Finally, while both systems cover all medically necessary benefits and the medical treatment quality is very similar, waiting times in the outpatient sector are shorter for the privately insured as outpatient reimbursement rates are structurally higher (Werbeck et al., 2021). Thus, the privately insured are more profitable for providers because sometimes PKV benefit packages are more generous than the GKV benefit package. As a consequence, they enjoy higher service quality (but potentially also overtreatment) by physicians.

The fact that some population subgroups have the right to choose between GKV and PKV has not only been criticized on the ground of fairness, but it implies selection into the PKV system. Specifically, as initial premiums are risk-rated, those who have pre-existing conditions and are unhealthy, likely decide to stay in the public system. As discussed, income, family status, and spouse's labor supply decisions also determine the decision to sign a GLTHI contract. Further, preferences and risk tolerance may determine the decision to opt out of GKV.

A1.1 Switching from PKV to GKV

There exist very limited institutional exemptions for PKV insured to return to the public GKV system. Below we provide empirical evidence on the switching rates.

First, for PKV insured above the age of 55, switching back to GKV is essentially impossible, even when their income decreases substantially, or when they become unemployed. For those above 55, one of the few options would be to exit the labor force and enroll under the GKV family plan of the spouse, if available. Rules for switching back to the GKV have been very strict for older employees. This is to avoid that individuals join the private system when young and healthy, and switch back to the public system when old and sick with little income (and thus low income-dependent contribution rates).

Second, PKV insured employees below the age of 55 can only return to the public GKV system if they become unemployed or if their gross wage from dependent employment permanently drops below the income threshold below which one is mandatorily insured with the GKV ("pflichtversichert").[A51] However, permanently switching to the GKV implies loosing the entire old-age provisions.

Third, PKV insured who are self-employed and below the age of 55 can only switch to GKV if they give up their business and become an employee with a gross salary below the income threshold

the choice between GKV and PKV. The other states only approve such requests in exceptional cases.

[A51] Assuming an average annual premium of €3,900 (as observed in our data), for an equally high GKV premium (15.5% of the gross wage), annual labor income would need to be as low as €25,000. Hence, artifically reducing income just to be able to join the GKV does not make sense for the overwhelming majority of cases.

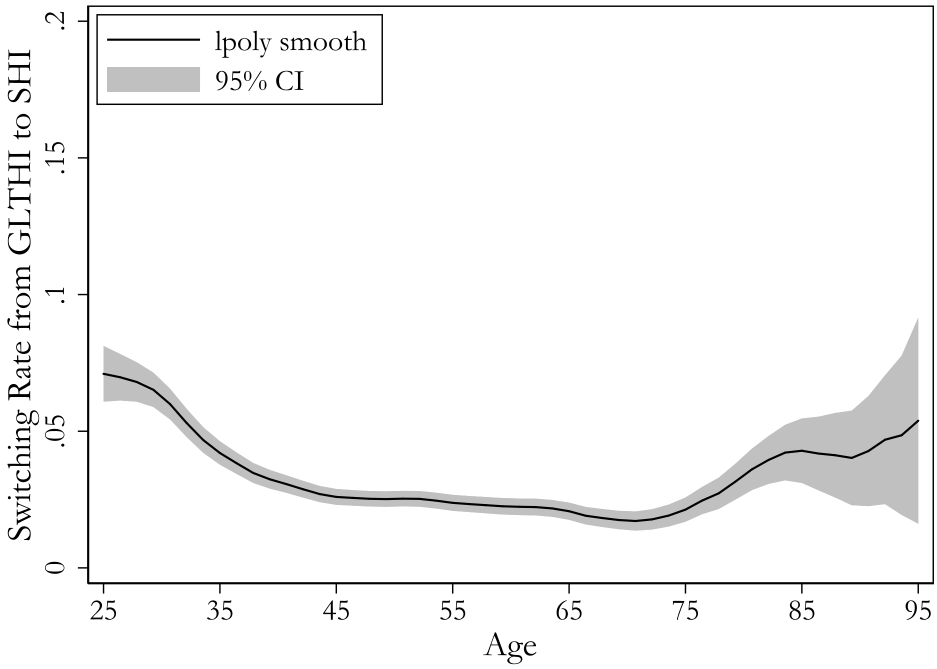


Figure A1: Likelihood to Return to GKV by Age

Notes: The figure shows the probability of returning from the private (PKV) to the public (GKV) system, by age, smoothed with an Epanechnikov kernel, degree 0, bandwidth 2.6. Source: Own calculations using SOEP (2018).

(see Social Code Book V, Para. 6 for details of the law, Büser, 2012; Cecu, 2018).

Data from the Association of German Private Healthcare Insurers (*Verband der Privaten Krankenversicherung*) show that 124,900 individuals, or 1.4 percent of all PKV insured, switched from PKV to GKV and 145,000 switched from GKV to PKV (Association of German Private Healthcare Insurers, 2022). Since 1997, the number of switchers to the GKV system has been very stable between 124,900 and 154.800 per year (Bundesministerium für Gesundheit, 2022). [A52] Figure A1 uses representative SOEP data to plot switching rates by age. As seen, the likelihood to return to GKV decreases substantially between the age of 25 and 35. We conjecture that this is mostly because those who were privately insured as students enter the labor market and have to enroll in GKV if their gross wages are below the mandatory income threshold. Switching rates remain stable at a low level between age 40 and age 75, and then slightly increase again. Using a fixed effects regression for the probability of switching to GKV among the universe of Germans who were at least once GLTHI policyholder, we find few significant predictors of switching back to GKV. In particular, health care utilization (num-

[A52]The total number of PKV insured increased from in 1991 to 8.976.400 in 2011. Since then, it has decreased slightly to 8.723.900 in 2020 (Bundesministerium für Gesundheit, 2022; Association of German Private Healthcare Insurers, 2022). Several reforms in the last decades have at least partly determined switching rates over time: The *Gesundheitsreformgesetz* of December 20, 1988 substantially tightened the possibility of joining the PKV for pensioners; the *Gesundheitsstrukturgesetz*, passed on December 21, 1992, introduced the free choice of GKV sickness funds, along with other provisions about the regulation of private insurers. Due to these and other reforms (e.g. the *GKV-Wettbewerbsstärkungsgesetz* of 2007), the GKV-PKV switching rate as a share of all GLTHI policyholders has declined over time.

Figure A2: Age Distribution of Initial Plan Inception

Notes: The figure shows the age distribution for individuals who first join the private insurance company. Source: German Claims Panel Data.

ber of hospital nights and doctor visits) are not significant predictors. The results of this analysis are available upon request.

A2 GLTHI Premium Calculation: Further Details

This section first summarizes the institutional details of the PKV lifecycle premium calculation. Hofmann and Browne (2013) and Atal et al. (2019) provide English references on the topic.

A2.1 Graphical Depiction of Basic Principles

Figure A3 illustrates the principle of constant lifecycle premiums (Rosenbrock, 2010; Hofmann and Browne, 2013). It uses our claims data as input and showcases four risk-age combinations at initial enrollment: low vs. high health risks, and initial enrollment at 30 vs. 50 years. This illustration assumes constant health risk types over the lifecycle.[A53] The low risk type (the "healthy") corresponds to a hypothetical individual with no pre-existing conditions; we denote her expected lifecycle health expenditures conditional on survival as $E(m|\text{surv, low})$. The high risk type (the "sick") corresponds to a hypothetical individual who has 50 percent higher expected health expenditures than the low risk type at each age. We denote her expected lifecycle health expenditures conditional on survival as $E(m|\text{surv, high})$. Note that $E(m|\text{surv, low})$ and $E(m|\text{surv, high})$ would also represent the actuarially fair premiums of short-term spot contracts by age, for low and high risk types, respectively. $P_{30,\text{low}}$ ($P_{30,\text{high}}$) are the GLTHI premiums for a low (high) risk type who first enrolls in a GLTHI plan at age 30. Similarly, $P_{50,\text{low}}$ and $P_{50,\text{high}}$ are the premiums if the corresponding type joins a GLTHI plan at age 50.

Figure A3 has the following important features: First, in theory, premiums are stable over individuals' lifecycles. Front-loading dampens age-driven premium increases via the legal requirement of building individual capital stocks through so-called old-age provisions (*Altersrückstellungen*). The capital stock is the cumulative difference between premiums and expected claims, plus returns on investment from the capital stock.[A54] Second, premiums are higher for policyholders who sign a GLTHI contract later in their lives. This is because health care expenditures increase with age, and those who join later in life have fewer years to build up old-age provisions.[A55] Third, because of the initial risk rating, high risk types pay higher premiums than low risk types throughout their lives. The following subsection discusses more institutional details about the premium calculation. It also

[A53]We assume constant health risks only for the purpose of illustrating the basic front-loading principle. Our empirical analysis below features evolving health risks which is fundamental: First, front-loading can dampen the reclassification risk. Second, evolving health risks imply that individuals who start unhealthy may lapse their contract if their health improves over time. This introduces (downwards) reclassification risk even if premiums are constant within a given contract. Also, we must consider lapsation when calculating equilibrium premiums.

[A54]In 2020, the average capital stock built through old-age provisions was €28,195 ($32,425) per enrollee (Association of German Private Healthcare Insurers, 2022).

[A55]This is not necessarily true when health changes over time. With stochastic health, the initial premium may start to decrease at very high ages as, over time, the need to front-load for future health shocks decreases (see Section 6.1.)

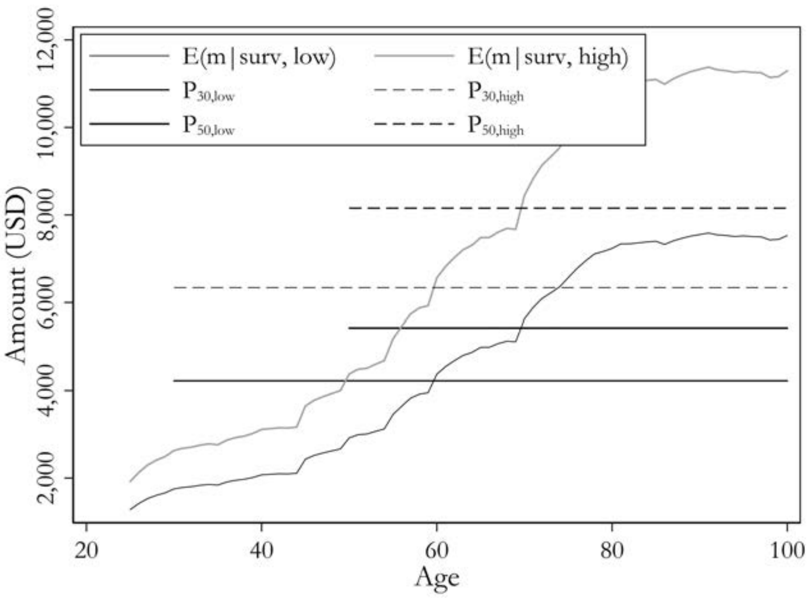


Figure A3: Illustration of GLTHI Lifecycle Premiums and Health Expenditures

Notes: The figure illustrates the principles of initial risk-rating and break-even in the life-cycle for GLTHI plans, and it is based on Rosenbrock (2010) and Hofmann and Browne (2013). Source: Own calculations using German Claims Panel Data.

discusses why, in reality, real premiums may not be flat over the lifecycle.

A2.2 Additional Details

The initial GLTHI premium is individually underwritten, but all subsequent premium changes over the lifecycle must be community rated.[A56] Premiums consist of several components. The *Kalkulationsverordnung (KalV)* regulates the actuarial calculations. These calculations have to be approved by a federal financial regulatory agency (the *Bundesanstalt für Finanzdienstleistungsaufsicht, BaFin*). The *KalV* specifies that premiums have to be a function of the expected per capita health care claims (*Kopfschäden*, which depend on the plan chosen, age, gender, and health risks),[A57] the assumed guaranteed interest rate (*Rechnungszins*), the probability to lapse (*Stornowahrscheinlichkeit*), and the life expectancy (*Sterbewahrscheinlichkeit*).

One important and distinct characteristic of the GLTHI market is the legal obligation of insurers to

[A56] The only exception is the Basic Plan (*'Basistarif*). All insurers have to offer a Basic Plan. It follows the standardized GKV plan with the same essential benefits and actuarial values. For the Basic Plan, guaranteed issue exists for people above 55 and those who joined the GLTHI after 2009. The maximum premium is capped at the maximum GKV contribution (2021: €769,16 per month). The legislature mandated the Basic Plan to provide an "affordable" private option for GLTHI enrollees who cannot switch back to GKV, are uninsured, would have to pay excessive premiums, or would be denied coverage. However, the demand for the Basic Plan has been negligible; thus henceforth, we will abstain from it. In 2019, in the entire GLTHI, only 32,400 people, or 0.4 percent, were enrolled in the Basic Plan (Association of German Private Healthcare Insurers, 2020). In our data, only 1,006 enrollees chose the basic plan in 2010.

[A57] Gender rating was allowed until December 21, 2012. After this date, for new contracts, all insurers in the European Union (EU) have to provide unisex premiums as the EU Court of Justice banned gender rating as discriminatory (Schmeiser et al., 2014) build up *old-age provisions*, typically until age 60 of the policyholder. Atal et al. (2019) provide further details on this institutional element. The old-age provisions accumulated early in the policyholder's lifecycle serve as the capital stock to cover higher health expenditures later in the lifecycle.

Premiums are then calculated under the basic principle of a constant lifecycle premium. This premium is supposed to be sufficient to cover the policyholder's lifecycle health care expenses. Section 3.1 formally expresses this principle. Thus, in young ages, premiums exceed expected claims while in old ages, premiums fall short of expected claims—a phenomenon known as "front-loading" in long-term insurance contracts (Hendel and Lizzeri, 2003; Nell and Rosenbrock, 2007, 2009; Fang and Kung, 2020).[A58]

While, theoretically, premiums are stable over individuals' lifecycles, in reality, nominal (and also real) premiums do increase. The main factors that trigger such "premium adjustments" (*Beitragsanpassungen*) are structural and unexpected changes in (i) life expectancy, (ii) health care consumption, (iii) health care prices, for example, due to new medical technology,[A59]and (iv) and economic fundamentals. An example of (iv) is the shift of central banks to a super-low interest rate environment over the past decade; this shift implies a significant decrease in the returns to risk-free capital investment. Because GLTHI insurers (like life insurers) are heavily invested in the bond market, premium adjustments are a consequence of super-low interest rates.[A60]

In some cases, premium adjustments are not only allowed, but *required* by the financial regulatory oversight agency *BaFin*. This is to ensure financial stability within the regulatory framework of the *Versicherungsvertragsgesetz (VVG)*, the *Versicherungsaufsichtsgesetz (VAG)*, and the *KalV*.[A61] Most insurers have to follow the *Solvency II* reporting requirements. Each year, insurers have to check whether their underlying assumptions to calculate premiums and old age provision are still accurate. If they deviate by a certain amount, they have to adjust the premiums, which can result in two-digit premium increases, bad press, and lawsuits (Krankenkassen-Zentrale (KKZ), 2020).[A62]

[A58] Such front-loading creates a "lock-in" effect, in addition to the lock-in induced by guaranteed renewability (Atal, 2019). To strengthen consumer power and reduce this lock-in, the German legislature made a standardized portion of these old-age provisions portable across insurers for contracts signed after Jan 1, 2009; see Atal et al. (2019) for an evaluation of this reform. For existing contracts, Atal et al. (2019) do not find a significant impact on the likelihood to switch insurers.

[A59]The Health Care Reform 2000 (*GKV-Gesundheitsreformgesetz 2000*) introduced a mandatory 10 percent premium surcharge up to age 60 to dampen structural increases in health care spending due to medical progress. This surcharge only applies to GLTHI contracts signed after January 1, 2000 (see article 14 of GKV-Gesundheitsreformgesetz (2000)).

[A60]The *KalV* has traditionally capped the assumed return on equity—the "guaranteed interest rate" (*Rechnungszins*)—at 3.5 percent for the premium calculation. This was the case for five decades. However, in 2016 for the first time, the average net return on investment has dropped below 3.5 percent, which is why the *German Actuary Association* has issued a new guideline to calculate the new insurer-specific "maximum allowed interest rate" (*Höchstrechnungszins*), see Deutsche Aktuarvereinigung (DAV) (2019).

[A61]Effective January 1, 2016 the *KalV* has been replaced by the *Krankenversicherungsaufsichtsverordnung (KVAV)*.

[A62]All premium adjustments have to be legally checked and approved by 16 independent actuaries who are appointed by the *BaFin*. However, some plaintiffs in lawsuits argue that some of these actuaries would not be sufficiently independent. Other reasons of courts to declare a premium increase as "not justified" were insufficient explanations by the insurers or a

However, average nominal premium increases have been moderate in an international comparison. According to data by the Association of German Private Healthcare Insurers (*Verband der Privaten Krankenversicherung*) total revenues from premiums increased annually between 0 and 4.5 percent from 2011 to 2020 (Bundesministerium für Gesundheit, 2022); the industry reports average annual premium increases of 2.6% from 2012-2022 (Bahnsen and Wild, 2021). Most important for our analysis is that, after the initial risk rating, premium adjustments do not depend on policyholders' evolving health status.

A3 Proof for Lemma 1

We prove the lemma through induction. With $b_T = 0$, we first show that Equation (3) holds for period $T-1$, i.e. $b_{T-1}(\xi_T|\Xi_{T-1})$. Then, we show that Equation (3) holds for period t, given it holds for $t+1$. First, note that for the securities to replicate the premium path of the GLTHI contract, the following equations must hold

$$-\tilde{P}_t(\Xi_t) = -E(m|\xi_t) - \sum_{\xi_{t+1}\in\mathcal{Z}} \delta\pi(\xi_{t+1} \mid \xi_t) b_t(\xi_{t+1}|\Xi_t) + b_{t-1}(\xi_t|\Xi_{t-1}) \quad \text{for } t \in \{1, 2, ..., T\} \tag{8}$$

with $b_0 = 0$ and $b_T = 0$. The previous expressions ensure that the addition of the premium of the short-term contract and the net proceeds from the Arrow securities equate the premium paid in the GLTHI contract after history Ξ_t.

Terminal Period. In the terminal period, the premium of a GLTHI contract in the spot market is equal to the premium of a short term contract: $P(\xi_T) = E(m_T \mid \xi_T)$. The premium paid under the GLTHI contract is therefore $\tilde{P}_T(\Xi_T) = \min\{P_T(\xi_T), \tilde{P}_{T-1}(\Xi_{T-1})\}$. There are two relevant cases:

1. $P_T(\xi_T) \leq \widetilde{P}_{T-1}(\Xi_{T-1})$ (*lapsation in* T). From equation (8), we get:

$$b_{T-1}(\xi_T \mid \Xi_{T-1}) = E(m_T \mid \xi_T) - \tilde{P}_T(\Xi_T) = P_T(\xi_T) - P_T(\xi_T) = 0$$

2. $P_T(\xi_T) > \widetilde{P}_{T-1}(\Xi_{T-1})$ (*no lapsation in* T). From equation (8), we get:

$$b_{T-1}(\xi_T \mid \Xi_{T-1}) = E(m_T \mid \xi_T) - \tilde{P}_T(\Xi_T) = E(m_T \mid \xi_T) - \tilde{P}_{T-1}(\Xi_{T-1})$$

deliberate initial underpricing of premiums in the first year to attract enrollees (Krankenkassen-Zentrale (KKZ), 2020).

Therefore equation (3) holds for $T-1$.

Period $t < T-1$. We now consider a period $t < T-1$. First, note that $\mathbb{1}\left(P_{t+2}(\xi_{t+2}) > \tilde{P}_{t+1}(\Xi_{t+1})\right) = 1 \iff b_{t+1}(\xi_{t+2}|\Xi_{t+1}) > 0$, and therefore,

$$\pi\left(\xi_{t+2} \mid \xi_{t+1}\right) b_{t+1}\left(\xi_{t+2} \mid \Xi_{t+1}\right) = \pi\left(\xi_{t+2} \mid \xi_{t+1}\right) b_{t+1}\left(\xi_{t+2} \mid \Xi_{t+1}\right) \mathbb{1}\left(P_{t+2}(\xi_{t+2}) > \tilde{P}_{t+1}(\Xi_{t+1})\right)$$

Given that equation (3) holds for period $t+1$,

$$\begin{aligned}
&\pi\left(\xi_{t+2} \mid \xi_{t+1}\right) b_{t+1}\left(\xi_{t+2} \mid \Xi_{t+1}\right) = \left(E\left(m_{t+2} \mid \xi_{t+2}\right) - \tilde{P}_{t+1}(\Xi_{t+1})\right) \pi\left(\xi_{t+2} \mid \xi_{t+1}\right) \mathbb{1}\left(P_{t+2}(\xi_{t+2}) > \tilde{P}_{t+1}(\Xi_{t+1})\right) \\
&+ \left(\sum_{\tau > t+2}^{T} \sum_{z \in \mathcal{Z}} \delta^{\tau-(t+2)} (E\left(m_\tau \mid \xi_\tau\right) - \tilde{P}_{t+1}(\Xi_{t+1})) \times q_\tau\left(z \mid \xi_{t+2}, \mathbf{P}_{t+3}^{\tau}, \tilde{P}_{t+1}\left(\Xi_{t+1}\right)\right)\right) \pi\left(\xi_{t+2} \mid \xi_{t+1}\right) \mathbb{1}\left(P_{t+2}(\xi_{t+2}) > \tilde{P}_{t+1}(\Xi_{t+1})\right)
\end{aligned}$$

By definition, $q_\tau\left(\xi_\tau \mid \xi_{t+1}, \mathbf{P}_{t+2}^{\tau}, \tilde{P}_t(\Xi_t)\right) = \prod_{j=t+2}^{\tau}\left[\mathbb{1}(P_j(\xi_j) > \tilde{P}_t(\Xi_t))\pi(\xi_{j+1}|\xi_j)\right]$. Therefore, we can re-write the expression above as:

$$\begin{aligned}
\pi\left(\xi_{t+2} \mid \Xi_{t+1}\right) b_{t+1}\left(\xi_{t+2} \mid \Xi_{t+1}\right) &= \left(E\left(m_{t+2} \mid \xi_{t+2}\right) - \tilde{P}_{t+1}(\Xi_{t+1})\right) \times q_{t+2}\left(\xi_{t+2} \mid \xi_{t+1}, \mathbf{P}_{t+2}^{t+2}, \tilde{P}_{t+1}\left(\Xi_{t+1}\right)\right) \\
&+ \sum_{\tau > t+2}^{T} \sum_{z \in \mathcal{Z}} \delta^{\tau-(t+2)} (E\left(m_\tau \mid \xi_\tau\right) - \tilde{P}_{t+1}(\Xi_{t+1})) \times q_\tau\left(z \mid \xi_{t+1}, \mathbf{P}_{t+2}^{\tau}, \tilde{P}_{t+1}\left(\Xi_{t+1}\right)\right)
\end{aligned}$$

Summing across ξ_{t+2} and multiplying by δ, we get:

$$\delta \sum_{\xi_{t+2}} \pi\left(\xi_{t+2} \mid \Xi_{t+1}\right) b\left(\xi_{t+2} \mid \Xi_{t+1}\right) = \sum_{\tau > t+1}^{T} \sum_{z \in \mathcal{Z}} \delta^{\tau-(t+1)} [E\left(m_\tau \mid \xi_\tau\right) - \tilde{P}_{t+1}(\Xi_{t+1})] \times q_\tau\left(z \mid \xi_{t+1}, \mathbf{P}_{t+2}^{\tau}, \tilde{P}_{t+1}\left(\Xi_{t+1}\right)\right)$$

We now consider the two relevant cases.

1. $P_{t+1}\left(\xi_{t+1}\right) \leq \widetilde{P}_t\left(\Xi_t\right)$ (*lapsation in $t+1$*). We have $\tilde{P}_{t+1}(\Xi_{t+1}) = P_{t+1}(\xi_{t+1})$. From equation (8) we get:

$$\begin{aligned}
&b_t\left(\xi_{t+1} \mid \Xi_t\right) = E\left(m_{t+1} \mid \xi_{t+1}\right) - \tilde{P}_{t+1}\left(\Xi_{t+1}\right) + \delta \sum_{\xi_{t+2}} \pi\left(\xi_{t+2} \mid \Xi_{t+1}\right) b_{t+1}\left(\xi_{t+2} \mid \Xi_{t+1}\right) \\
&= E\left(m_{t+1} \mid \xi_{t+1}\right) - \tilde{P}_{t+1}\left(\Xi_{t+1}\right) + \sum_{\tau > t+1}^{T} \sum_{z \in \mathcal{Z}} \delta^{\tau-(t+1)} [E\left(m_\tau \mid \xi_\tau\right) - \tilde{P}_{t+1}(\Xi_{t+1})] \times q_\tau\left(z \mid \xi_{t+1}, \mathbf{P}_{t+2}^{\tau}, \tilde{P}_{t+1}\left(\Xi_{t+1}\right)\right) \\
&= E\left(m_{t+1} \mid \xi_{t+1}\right) - P_{t+1}\left(\xi_{t+1}\right) + \sum_{\tau > t+1}^{T} \sum_{z \in \mathcal{Z}} \delta^{\tau-(t+1)} [E\left(m_\tau \mid \xi_\tau\right) - P_{t+1}(\xi_{t+1})] \times q_\tau\left(z \mid \xi_{t+1}, \mathbf{P}_{t+2}^{\tau}, P_{t+1}\left(\xi_{t+1}\right)\right) \\
&= 0
\end{aligned}$$

where the last step follows from the insurer's zero-profit condition.

2. $P_{t+1}(\xi_{t+1}) > \widetilde{P}_t(\Xi_t)$ (*no lapsation in* $t+1$). We have $\tilde{P}_{t+1}(\Xi_{t+1}) = \widetilde{P}_t(\Xi_t)$ .From equation (8) we get:

$$
\begin{aligned}
& b_t(\xi_{t+1} \mid \Xi_t) = E(m_{t+1} \mid \xi_{t+1}) - \tilde{P}_{t+1}(\Xi_{t+1}) + \\
& \sum_{\tau > t+1}^{T} \sum_{z \in \mathcal{Z}} \delta^{\tau-(t+1)}[E(m_\tau \mid \xi_\tau) - \tilde{P}_{t+1}(\Xi_{t+1})] \times q_\tau\left(z \mid \xi_{t+1}, \mathbf{P}_{t+2}^{\tau}, \tilde{P}_{t+1}(\Xi_{t+1})\right) \\
& = E(m_{t+1} \mid \xi_{t+1}) - \tilde{P}_t(\Xi_t) + \sum_{\tau > t+1}^{T} \sum_{z \in \mathcal{Z}} \delta^{\tau-(t+1)}[E(m_\tau \mid \xi_\tau) - \tilde{P}_t(\Xi_t)] \times q_\tau\left(z \mid \xi_{t+1}, \mathbf{P}_{t+2}^{\tau}, \tilde{P}_t(\Xi_t)\right)
\end{aligned}
$$

Therefore, equation (3) also holds for period t.

Then the equation holds for all $t \in \{1, 2, ..., T\}$.

A3.1 Characterizing Optimal Contracts with Arrow Securities

This section shows how to implement the optimal dynamic long-term contracts as in Ghili et al. (2022) with one-sided commitment by trading state-contingent one-period Arrow securities.

First, we define the individual's consumption when only short-term contracts are available $c_t(\xi_t)$:

$$
c_t(\xi_t) = y_t - E(m_t \mid \xi_t).
$$

Equation (2) provides the individual's long-term guaranteed consumption offered by a competitive insurer. We define the guaranteed consumption from the previous period, $\tilde{c}_t$:

$$
\tilde{c}_t(\Xi_t) = \max\left\{\widetilde{c}_{t-1}(\Xi_{t-1}), \quad \overline{c}_t(\xi_t)\right\}
$$

We aim to find the quantities of Arrow securities that would replicate the consumption path under the optimal contract when combined with short-term contracts. In particular, it must be true that

$$
\tilde{c}_t(\Xi_t) = y_t - E(m|\xi_t) - \sum_{\xi_{t+1} \in \mathcal{Z}} \delta\pi(\xi_{t+1} \mid \xi_t) b_t(\xi_{t+1}|\Xi_t) + b_{t-1}(\xi_t|\Xi_{t-1}) \quad \text{for } t \in \{1, 2, ..., T\}
$$

with $b_0 = 0$ and $b_T = 0$

Lemma 2 *The consumption path under the optimal dynamic contract can be replicated by purchasing short-term insurance contracts supplemented by Arrow securities. The quantity of Arrow securities bought after history Ξ_t that pay one dollar in state ξ_{t+1} are equal to*

$$b_t\left(\xi_{t+1} \mid \Xi_t\right) = \begin{cases} 0 \quad \text{if } \bar{c}_{t+1}\left(\xi_{t+1}\right) > \tilde{c}_t\left(\Xi_t\right) \\ (\tilde{c}_t(\Xi_t) - c_{t+1}(\xi_{t+1})) + \\ \quad \sum_{\tau>t+1}^{T} \sum_{z\in\mathcal{Z}} \delta^{\tau-(t+1)}[\tilde{c}_t(\Xi_t) - c_\tau(\xi_\tau)] q_\tau\left(z \mid \xi_{t+1}, \mathbf{C}_{t+2}^{\tau}, \tilde{c}_t(\Xi_t)\right) \quad \text{otherwise} \end{cases} \tag{9}$$

where $q_\tau\left(z \mid \xi_{t+1}, \mathbf{C}_{t+2}^{\tau}, \tilde{c}_t(\Xi_t)\right)$ is the probability that (i) $\xi_\tau = z$, and (ii) the enrollee does not lapse (or die) between periods $t+1$ and τ, given the subsequent equilibrium consumption $\mathbf{C}_{t+2}^{\tau}$ and the guaranteed-renewable consumption $\tilde{c}_t(\Xi_t)$.

A4 Descriptive Statistics for GLTHI Dataset and SOEP Dataset

Table A1: Summary Statistics: German Claims Panel Data

	Mean	SD	Min	Max	N
Socio-Demographics					
Age (in years)	45.5	11.4	25.0	99.0	1,867,465
Female	0.276	0.447	0.0	1.0	1,867,465
Policyholder since (years)	6.5	5.0	1.0	40.0	1,867,465
Client since (years)	12.8	11.0	1.0	86.0	1,867,465
Employee	0.336	0.473	0.0	1.0	1,867,465
Self-Employed	0.486	0.500	0.0	1.0	1,867,465
Civil Servant	0.132	0.338	0.0	1.0	1,867,465
Health Risk Penalty	0.358	0.480	0.0	1.0	1,867,465
Pre-Existing Condition Exempt	0.016	0.126	0.0	1.0	1,867,465
Health Plan Parameters					
TOP Plan	0.377	0.485	0.0	1.0	1,867,465
PLUS Plan	0.338	0.473	0.0	1.0	1,867,465
ECO Plan	0.285	0.451	0.0	1.0	1,867,465
Annual premium (USD)	4,749	2,157	0	33,037	1,867,318
Annual risk penalty (USD)	157	453	0	21,752	1,867,465
Deductible(USD)	675	659	0	3,224	1,867,465
Total Claims (USD)	3,289	8,577	0	2,345,126	1,867,465

Source: German Claims Panel Data. *Policyholder since* is the number of years since the client has enrolled in the current plan; *Client since* is the number of years since the client joined the company. *Employee* and *Self-Employed* are dummies for the policyholders' current occupation. *Health Risk Penalty* is a dummy that is one if the initial underwriting led to a health-related risk penalty on top of the factors age, gender, and type of plan; *Pre-Existing Conditions Exempt* is a dummy that is one if the initial underwriting led to exclusions of pre-existing conditions. The mutually exclusive dummies *TOP Plan*, *PLUS Plan* and *ECO Plan* capture the generosity of the plan. *Annual premium* is the annual premium, and *Annual Risk Penalty* is the amount of the health risk penalty charged. *Deductible* is the deductible and *Total Claims* the sum of all claims in a calendar year. See Section 4.1 for further details.

Table A2: Summary Statistics: German Socio-Economic Panel Study

	Mean	SD	Min	Max	N
Socio-Demographics					
Female	0.5217	0.4995	0	1	530,228
Age	46.9119	17.4922	17	105	530,228
No degree yet	0.058	0.2338	0	1	530,228
Dropout of high school	0.0378	0.1908	0	1	530,228
Degree after 8/9 years of schooling (Ed 8)	0.3619	0.4805	0	1	530,228
Degree after 10 years of schooling (Ed 10)	0.2737	0.4459	0	1	530,228
Degree after 13 years of schooling (Ed 13)	0.1746	0.3796	0	1	530,228
Employment					
Civil servant	0.0393	0.1943	0	1	530,228
Self-employed	0.0624	0.2419	0	1	530,228
White collar	0.2736	0.4458	0	1	530,228
Full-time employed	0.4152	0.4928	0	1	530,228
Part-time employed	0.1402	0.3471	0	1	530,228
Income Measures in 2016 USD					
Monthly gross wage	2,940	2,506	0	215,093	310,460
Monthly net wage	1,921	1,527	0	134511.5	310,460
Individual annual total income	20,361	24,434	0	2,580,000	530,228
Equivalized post-tax post-transfer annual income	26,433	18,731	0	2,155,394	530,228
Insurance and Utilization					
Hospital nights in past calendar year	1.6652	8.3794	0	365	530,228
Doctor visits in past 3 months	2.4941	4.1436	0	99	461,971
Privately insured	1	0	1	1	57,558

Source: SOEP (2018), the long version from 1984 to 2016. Whenever the number of person-year observations is less than 530,228 the question was not asked in all years from 1984 to 2016. For example, *Doctor visits in past 3 months* has only been routinely asked since 1995. *Privately insured* indicates that 57,558/530,228=10.8% of all observations are by people who are insured on the GLTHI market. All income measures have been consistently generated and cleaned by the SOEP team; e.g., *Monthly gross wage* is labeled *labgro* and *Monthly net wage* is labeled *labnet* in SOEP (2018). See Section 4.2 for a detailed discussion of the variables.

A5 Risk Classification: Further Details and Robustness Checks

A5.1 Efficient Classification

According to the actuarial science literature (Finger, 2001), an efficient risk classification system has two properties: *homogeneity*—meaning that individuals in the same risk category have similar risk and *separation*—meaning that the categories have sufficiently different expected claims to justify distinct categories.[A63] We show in Appendix A5 that applying k-means clustering to

[A63]For instance, in Figure 1, it is easy to see that equally-sized categories are unlikely to be optimal as they would assign similar individuals in terms of λ^* into different categories in the left tail of the distribution, failing the *separation* principle.

We define a *risk classification* as a surjective function $f_K : \Re^n_+ \to \{\lambda \in \mathbb{Z} : 1 \leq \lambda \leq K\}$, where $\Re^n_+$ is the state space (i.e. λ^*_t and its $n-1$ lags). Denote this classification function $\lambda_t = f_K(\Lambda^*_t(n))$, where $\lambda_t \in \{1, \ldots, K\}$ is the risk category assigned to a person with those ACG scores. The *efficient* risk classification f_K maximizes the "structure variance"

$$SV(f_K) = \mathrm{Var}(m_t) - \sum_{k=1}^{K} \Pr(\lambda_t = k) \mathrm{Var}(m_t \mid \lambda_t = k), \tag{10}$$

where m_t is individual annual health expenditures (Finger, 2001). The structure variance $SV(f_K)$ is thus the total variance less the weighted sum of within-class variances of health expenditures. Put differently, the efficient classification maximizes the variance of mean expenditure across groups. Applying the law of total variance to both terms in Equation (10), we can write the structure variance as:[A64]

$$SV(f_K) = \mathrm{Var}(\mathbb{E}(m_t \mid \Lambda^*_t(n))) - \sum_{k=1}^{K} \Pr(\lambda_t = k) \mathrm{Var}(\mathbb{E}(m_t \mid \Lambda^*_t(n)) \mid \lambda_t = k). \tag{11}$$

Note that the first term in Equation (11) is independent of the classification (as it is independent of the classes λ_t). Thus for a given K, finding the efficient classification system is equivalent to finding the classes λ_t that *minimize* the heterogeneity in expected expenditure within risk classes: $\sum_{k=1}^{K} \Pr(\lambda_t = k) \mathrm{Var}(\mathbb{E}(m_t \mid \Lambda^*_t(n)) \mid \lambda_t = k)$.

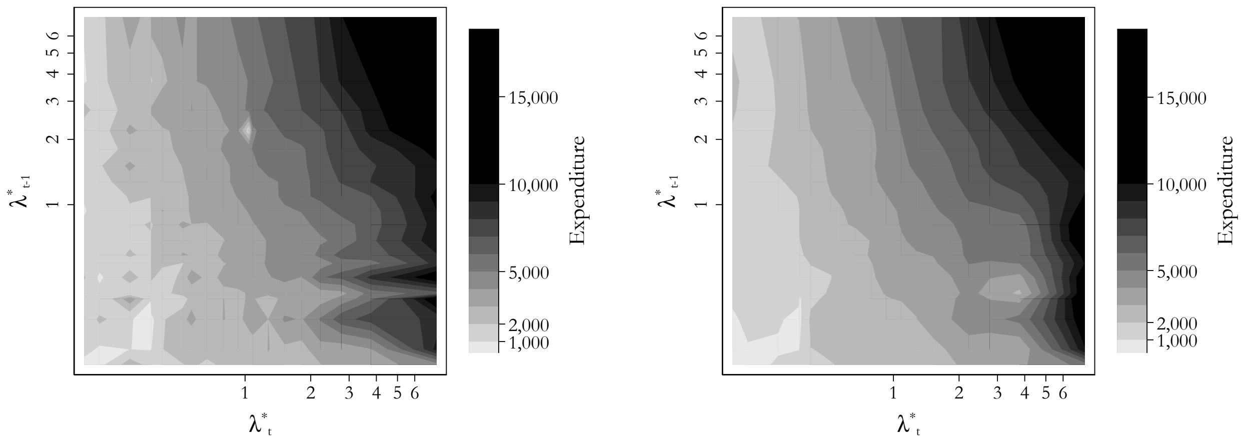
Three additional things are worth noting about Equation (11). First, only mean expenditures conditional on ACG scores $\mathbb{E}(m_t \mid \Lambda^*_t(n))$ matter for the classification system, whereas the dispersion of m_t around this mean is inconsequential. Second, minimizing heterogeneity within classes is incidentally what the k-means clustering method does (Lloyd, 1982; Athey and Imbens, 2019). Thus, we will apply k-means clustering of $\mathbb{E}(m_t \mid \Lambda^*_t(n))$ to determine the efficient classification system. Third, this implies that the efficient classification also maximizes the coefficient of determination (R^2) in a regression of expenditures on risk class indicators (Kriegel et al., 2017).

A5.2 Further Details

Smoothing. To get accurate predictions for expected claims along the entire distribution of risk, including the tails, we use cubic regression splines in the estimation of $\mathbb{E}(m_t \mid \Lambda^*_t(n))$. Figure A4 provides a comparison of mean expenditure by $\Lambda^*_t(n)$ before and after smoothing for $n = 2$.

In addition, it would assign individuals with substantial λ^* differences into identical categories in the right tail of the distribution, failing the *homogeneity* principle.

[A64]The law of total variance implies $\mathrm{Var}(m_t) = \mathbb{E}(\mathrm{Var}(m_t \mid \Lambda^*_t(n))) + \mathrm{Var}(\mathbb{E}(m_t \mid \Lambda^*_t(n)))$ and $\mathrm{Var}(m_t \mid \lambda_t = k) = \mathbb{E}(\mathrm{Var}(m_t \mid \Lambda^*_t(n)) \mid \lambda_t = k) + \mathrm{Var}(\mathbb{E}(m_t \mid \Lambda^*_t(n)) \mid \lambda_t = k)$.



(a) Raw Averages

(b) Smoothed Expenditure

Figure A4: Mean Expenditure by Λ_t^*.

Note: The left figure is based on average expenditure within each of 400 cells (ventiles in λ_t^* and λ_{t-1}^*). The right figure uses predicted values from a cubic spline regression. Source: German Claims Panel Data.

Stochastic Dominance. In their characterization of the optimal contract, Ghili et al. (2022) invoke an assumption of stochastic dominance. It requires that transition rates between risk categories—which are represented by the cumulative distribution function $F(\lambda_{t+1} \mid \lambda_t)$—satisfy first-order stochastic dominance in the following sense: if $\lambda_t' > \lambda_t$, then $F(\lambda_{t+1} \mid \lambda_t') \succ_{FSD} F(\lambda_{t+1} \mid \lambda_t)$. Figure A5 shows that this property holds for all pairwise combinations of (λ_t, λ_t') such that $\lambda_t' > \lambda_t$.

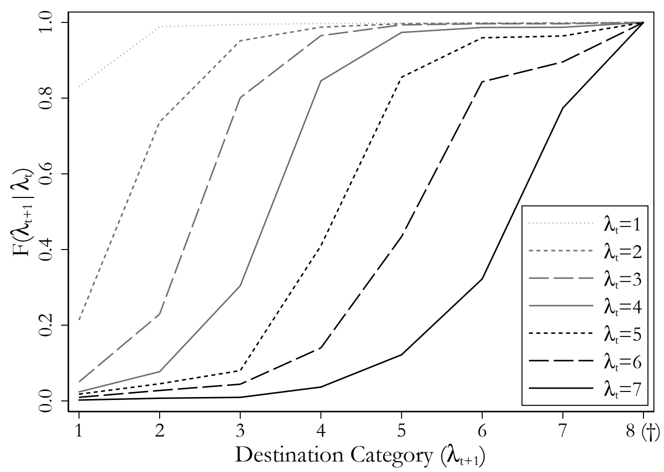


Figure A5: Stochastic Dominance.

Note: The figure shows the cumulative distribution function $F(\lambda_{t+1} \mid \lambda_t)$ for all pairwise combinations of (λ_t, λ_t') such that $\lambda_t' > \lambda_t$. Source: German Claims Panel Data.

Predicted Expenditures. Figure A6 shows the estimated mean expenditure by age for each risk category.

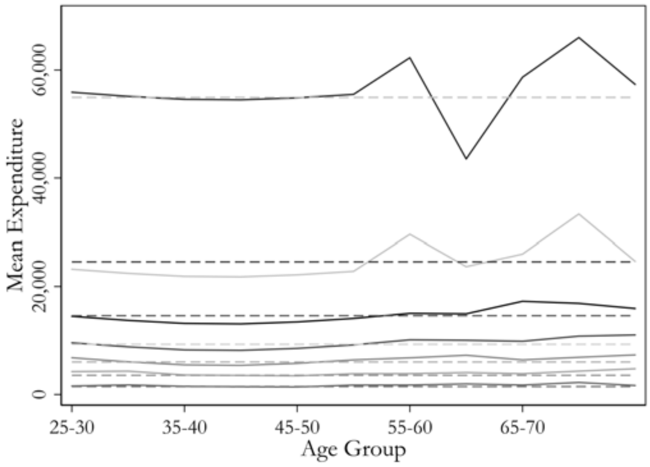


Figure A6: Predicted Health Expenditure

Note: Solid curves represent mean expenditure by age for each risk category λ_t, estimated according to Equation (5) in Section 5.2. The dashed lines represent the corresponding predictions assuming expenditure does not depend on age. Source: German Claims Panel Data.

Transition Matrices. Table A3 and A4 show risk category transition matrices for those aged 25-54 and 55+, respectively.

Table A3: λ Risk Category Transitions by Age Group—Ages 25–54

		λ_{t+1}							
Age	λ_t	1	2	3	4	5	6	7	8 (†)
25-29	1	0.8907	0.1024	0.0047	0.0011	0.0004	0.0003	0.0001	0.0004
	2	0.3197	0.4257	0.2020	0.0432	0.0077	0.0011	0.0003	0.0003
	3	0.1242	0.2829	0.4104	0.1404	0.0378	0.0043	0.0000	0.0000
	4	0.0892	0.1688	0.2484	0.3917	0.0860	0.0159	0.0000	0.0000
	5	0.0938	0.1250	0.0625	0.3750	0.2917	0.0521	0.0000	0.0000
	6	0.0909	0.0000	0.0455	0.2273	0.3182	0.3182	0.0000	0.0000
	7	0.0000	0.0000	0.0002	0.0045	0.0240	0.1447	0.7619	0.0647
30-34	1	0.8767	0.1145	0.0055	0.0018	0.0009	0.0002	0.0001	0.0003
	2	0.3212	0.4347	0.1909	0.0438	0.0080	0.0006	0.0001	0.0007
	3	0.1241	0.3015	0.4080	0.1409	0.0229	0.0016	0.0000	0.0011
	4	0.1039	0.1640	0.2407	0.3739	0.1032	0.0115	0.0007	0.0021
	5	0.0734	0.0911	0.0506	0.2911	0.3747	0.1089	0.0025	0.0076
	6	0.0422	0.0438	0.0529	0.1678	0.3628	0.2450	0.0525	0.0329
	7	0.0128	0.0115	0.0083	0.0574	0.1545	0.1663	0.4524	0.1368
35-39	1	0.8427	0.1480	0.0055	0.0022	0.0009	0.0002	0.0001	0.0004
	2	0.2798	0.4635	0.2113	0.0360	0.0076	0.0013	0.0000	0.0005
	3	0.1177	0.2379	0.4850	0.1288	0.0275	0.0028	0.0001	0.0002
	4	0.0719	0.0967	0.3055	0.4085	0.0999	0.0158	0.0003	0.0014
	5	0.0743	0.0493	0.0691	0.3402	0.3629	0.0958	0.0039	0.0045
	6	0.0415	0.0331	0.0340	0.1180	0.2958	0.4009	0.0455	0.0312
	7	0.0127	0.0088	0.0054	0.0409	0.1276	0.2757	0.3975	0.1313
40-44	1	0.8514	0.1392	0.0050	0.0024	0.0010	0.0003	0.0001	0.0006
	2	0.2862	0.4666	0.2050	0.0329	0.0075	0.0014	0.0001	0.0003
	3	0.1137	0.2229	0.5134	0.1225	0.0241	0.0022	0.0001	0.0011
	4	0.0790	0.0769	0.2936	0.4213	0.1113	0.0157	0.0003	0.0018
	5	0.0640	0.0392	0.0759	0.3281	0.3763	0.1055	0.0038	0.0072
	6	0.0295	0.0382	0.0342	0.1605	0.2773	0.3613	0.0539	0.0450
	7	0.0081	0.0091	0.0049	0.0502	0.1079	0.2240	0.4247	0.1710
45-49	1	0.8148	0.1736	0.0059	0.0028	0.0012	0.0006	0.0002	0.0009
	2	0.2267	0.5059	0.2229	0.0329	0.0093	0.0013	0.0001	0.0010
	3	0.0653	0.2027	0.5708	0.1309	0.0258	0.0031	0.0001	0.0012
	4	0.0427	0.0712	0.2877	0.4655	0.1153	0.0140	0.0005	0.0029
	5	0.0303	0.0438	0.0475	0.3570	0.3964	0.1101	0.0058	0.0090
	6	0.0153	0.0266	0.0211	0.1118	0.2919	0.4163	0.0607	0.0563
	7	0.0038	0.0057	0.0027	0.0314	0.1021	0.2321	0.4298	0.1923
50-54	1	0.8117	0.1740	0.0056	0.0035	0.0020	0.0008	0.0004	0.0020
	2	0.2283	0.4979	0.2228	0.0377	0.0101	0.0016	0.0002	0.0015
	3	0.0602	0.1799	0.5727	0.1509	0.0317	0.0027	0.0001	0.0018
	4	0.0398	0.0648	0.2660	0.4930	0.1160	0.0155	0.0007	0.0041
	5	0.0274	0.0387	0.0426	0.3666	0.3866	0.1182	0.0075	0.0124
	6	0.0130	0.0222	0.0179	0.1084	0.2688	0.4220	0.0746	0.0732
	7	0.0028	0.0042	0.0020	0.0265	0.0819	0.2049	0.4600	0.2176

Note: The table shows transition probabilities across risk categories by age group. Source: German Claims Panel Data.

Table A4: λ Risk Category Transitions by Age Group—Ages 55+

Age	λ_t	λ_{t+1} 1	2	3	4	5	6	7	8 (†)
55-59	1	0.7261	0.2537	0.0101	0.0037	0.0020	0.0013	0.0004	0.0027
	2	0.0932	0.6432	0.2123	0.0357	0.0110	0.0018	0.0004	0.0025
	3	0.0002	0.1739	0.6167	0.1690	0.0335	0.0044	0.0001	0.0024
	4	0.0001	0.0637	0.2426	0.5404	0.1287	0.0180	0.0007	0.0058
	5	0.0001	0.0356	0.0363	0.3758	0.4009	0.1282	0.0069	0.0163
	6	0.0000	0.0195	0.0145	0.1061	0.2662	0.4370	0.0650	0.0917
	7	0.0000	0.0037	0.0016	0.0260	0.0813	0.2126	0.4016	0.2732
60-64	1	0.7558	0.2147	0.0145	0.0044	0.0042	0.0019	0.0011	0.0033
	2	0.1023	0.6414	0.1981	0.0387	0.0120	0.0031	0.0004	0.0040
	3	0.0002	0.1612	0.6076	0.1836	0.0394	0.0053	0.0001	0.0028
	4	0.0001	0.0555	0.2243	0.5507	0.1419	0.0204	0.0008	0.0063
	5	0.0001	0.0292	0.0317	0.3610	0.4168	0.1370	0.0075	0.0168
	6	0.0000	0.0153	0.0122	0.0980	0.2660	0.4489	0.0686	0.0910
	7	0.0000	0.0028	0.0013	0.0235	0.0794	0.2136	0.4143	0.2651
65-69	1	0.3707	0.5949	0.0172	0.0076	0.0030	0.0015	0.0009	0.0042
	2	0.0624	0.6492	0.2407	0.0352	0.0065	0.0012	0.0004	0.0045
	3	0.0008	0.1058	0.6561	0.2082	0.0223	0.0013	0.0000	0.0056
	4	0.0002	0.0335	0.2013	0.6242	0.1261	0.0052	0.0005	0.0090
	5	0.0000	0.0128	0.0159	0.3546	0.4985	0.0763	0.0019	0.0400
	6	0.0000	0.0000	0.0107	0.0551	0.4067	0.3517	0.0195	0.1563
	7	0.0006	0.0066	0.0029	0.0264	0.0553	0.1690	0.5289	0.2103
70-74	1	0.3848	0.5793	0.0225	0.0060	0.0011	0.0003	0.0014	0.0048
	2	0.0070	0.6771	0.2554	0.0406	0.0105	0.0012	0.0000	0.0082
	3	0.0001	0.0810	0.6277	0.2599	0.0230	0.0014	0.0001	0.0068
	4	0.0002	0.0115	0.1625	0.6579	0.1404	0.0080	0.0002	0.0195
	5	0.0000	0.0015	0.0184	0.2829	0.5654	0.0736	0.0010	0.0572
	6	0.0000	0.0000	0.0000	0.0327	0.3039	0.4052	0.0065	0.2516
	7	0.0005	0.0056	0.0033	0.0184	0.0172	0.0263	0.7192	0.2094
75+	1	0.1770	0.5900	0.0442	0.0995	0.0598	0.0063	0.0083	0.0150
	2	0.0006	0.6237	0.2903	0.0471	0.0094	0.0012	0.0000	0.0277
	3	0.0000	0.0525	0.5876	0.2988	0.0254	0.0012	0.0000	0.0344
	4	0.0000	0.0029	0.1012	0.6668	0.1623	0.0055	0.0008	0.0605
	5	0.0000	0.0000	0.0060	0.2262	0.5581	0.0837	0.0028	0.1232
	6	0.0000	0.0000	0.0019	0.0206	0.3127	0.4064	0.0225	0.2360
	7	0.0000	0.0000	0.0000	0.0000	0.1111	0.1481	0.4630	0.2778

Notes: The table shows transition probabilities across risk categories by age group. Source: German Claims Panel Data.

A5.3 Robustness Checks

Winsorizing. First, we analyse the extent to which results are driven by outliers in m_{it}. It is of course desirable that outliers are considered in the classification, given their disproportionate contributions to means and variances; however, if the performance of the classification were widely different when they are not considered, it would cast doubt on how well the scheme performs with regard to less extreme risks. Therefore, we compared the performance of different classification schemes after the top percentile of expenditure had been been winsorized. Results are provided in Figure A7. As

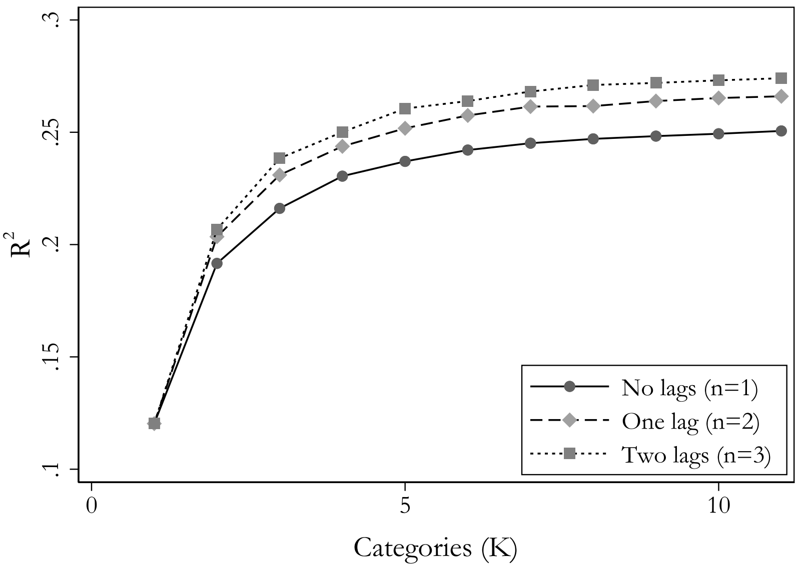


Figure A7: Performance of Alternative Risk Classifications: Winsorized Expenditure.

Note: The figure displays *unadjusted* R^2 of a regression of expenditures on risk category indicators. All results are robust to using the *adjusted* R^2. Each specification includes 21 age-gender fixed effects, year fixed effects and 79 plan fixed effects. Source: Own Calculations using German Claims Panel Data.

expected, the topcoding of outliers improves the predictive power of all schemes; however, their relative performance is unaffected by this change.

Lags of classes. Second, we compare two different ways of including a longer history of claims. Instead of expanding on the information set Λ_t before discretizing, we consider an alternative based on $\Lambda_t^* = \lambda_t^*$ but with the predictive power of the classification scheme interacted with its lags (i.e. a classification based on K^2 classes). Figure A8 provides the results, comparing the two alternatives $q = 0$ and $q = 1$ from above. In an interacted version, the classification uses $q = 0$ but is interacted with its lags in the regressions (leading effectively to K^2 classes). This interacted alternative has similar, even better, predictive power than $q = 1$. However, the variant with $q = 1$ achieves similar performance with a much smaller number of classes.

Sample selection. The results in Figure 2 are based on a sample of individuals who are observed over 4 years, since three lags are needed in Λ_{it}^*. In Figure A9 we check how robust the finding is to varying the observation window for sample selection. Sample 1 requires only that m_i and λ_t^* are observed. Sample 2 requires also that λ_{t-1}^* is observed, and sample 3 requires in addition that λ_{t-2}^* is observed. Figure A9 shows the result and that the predictive performance is sensitive to the sample used; however, the relative performance between schemes is the same regardless of the sample.

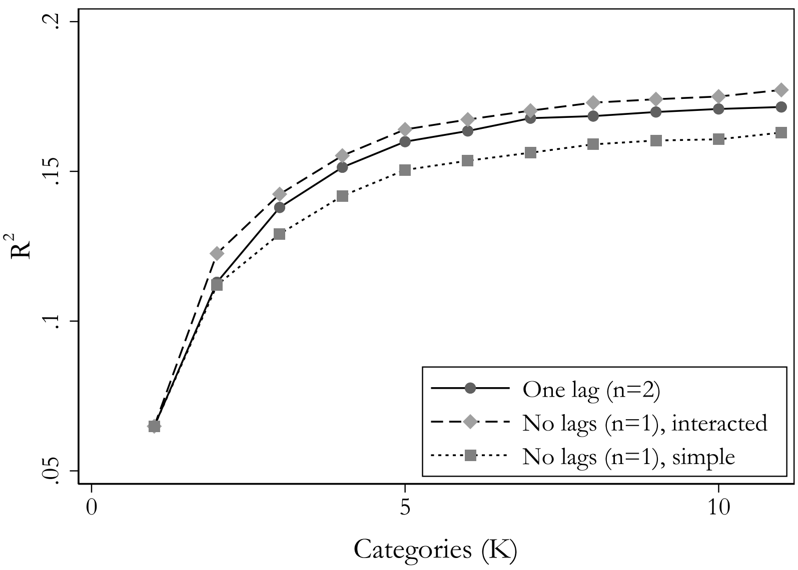


Figure A8: Performance of Alternative Risk Classifications: lags of classification.

Note: The figure displays *unadjusted* R^2 of a regression of expenditures on risk category indicators. All results are robust to using the *adjusted* R^2. Each specification includes 21 age-gender fixed effects, year fixed effects and 79 plan fixed effects. Source: Own Calculations using German Claims Panel Data..

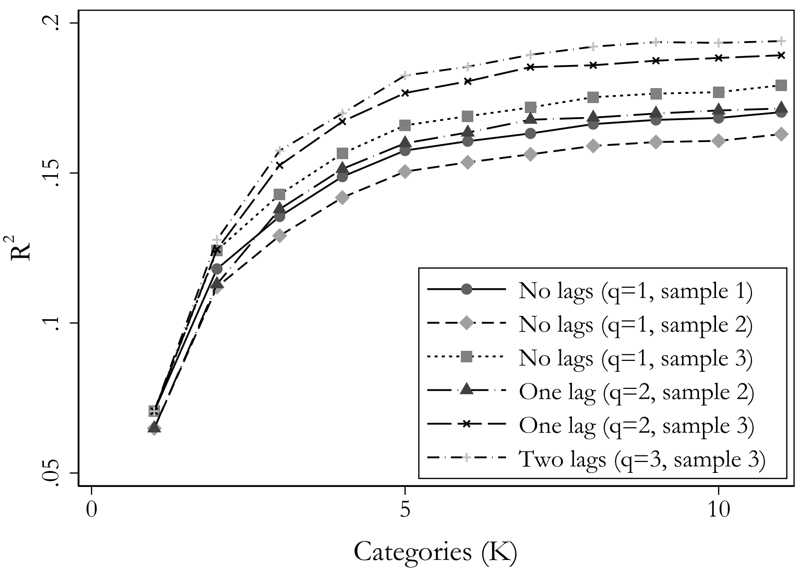


Figure A9: Performance of Alternative Risk Classifications: Different Samples.

Note: Each specification includes 21 age$\times$gender fixed effects, year fixed effects and 79 plan fixed effects. Source: German Claims Panel Data.

Sample with low deductible. This robustness section focuses on plans with low deductibles. We consider a stricter sample selection rule, where we only include plans with deductibles below \$400.[A65] These plans have approximately full coverage and thus more reliable information on the universe of health care expenditures. Summary statistics for this subsample are provided in Table A5. A comparison with the numbers in Table A1 makes clear that the two samples are very similar in terms of age, gender and history with the company. On the other hand, the restricted sample has a greater share of employees and civil servants, but a smaller share of self-employed. The plan characteristics are also similar to a great extent—with the obvious exceptions of deductible size and average claims.

Table A5: Summary Statistics: Low-Deductible Plans

	Mean	SD	Min	Max	N
Socio-Demographics					
Age (in years)	44.8	11.8	25.0	99.0	879,468
Female	0.256	0.437	0.0	1.0	879,468
Policyholder since (years)	7.7	5.3	1.0	40.0	879,468
Client since (years)	13.9	11.7	1.0	84.0	879,468
Employee	0.414	0.493	0.0	1.0	879,468
Self-Employed	0.281	0.449	0.0	1.0	879,468
Civil Servant	0.280	0.449	0.0	1.0	879,468
Health Risk Penalty	0.338	0.473	0.0	1.0	879,468
Pre-Existing Condition Exempt	0.015	0.121	0.0	1.0	879,468
Health Plan Parameters					
TOP Plan	0.342	0.475	0.0	1.0	879,468
PLUS Plan	0.397	0.489	0.0	1.0	879,468
ECO Plan	0.261	0.439	0.0	1.0	879,468
Annual premium (USD)	5,208	2,005	0	33,037	879,374
Annual risk penalty (USD)	133	347	0	21,214	879,468
Deductible(USD)	154	164	0	395	879,468
Total Claims (USD)	3,868	9,064	0	2,345,126	879,468

Source: German Claims Panel Data. *Policyholder since* is the number of years since the client has enrolled in the current plan; *Client since* is the number of years since the client joined the company. *Employee* and *Self-Employed* are dummies for the policyholders' current occupation. *Health Risk Penalty* is a dummy that is one if the initial underwriting led to a health-related risk add-on premium on top of the factors age, gender, and plan; *Pre-Existing Conditions Exempt* is a dummy which equals one if the initial underwriting led to a coverage exclusion of services for some conditions. The mutually exclusive dummies *TOP Plan*, *PLUS Plan* and *ECO Plan* capture the generosity of the plan. *Annual premium* is the annual premium, and *Annual Risk Penalty* is the amount of the health risk penalty charged. *Deductible* is the deductible and *Total Claims* the sum all claims in a calendar year. See Section 4.1 for further details.

Figure A10 compares the distributions of λ^* in the two samples. As expected, the zero-deductible

[A65] This is the lowest cutoff for the deductible which gives us a sufficient number of observations to analyze health risk transitions within each age group.

plans have higher ACG© scores.

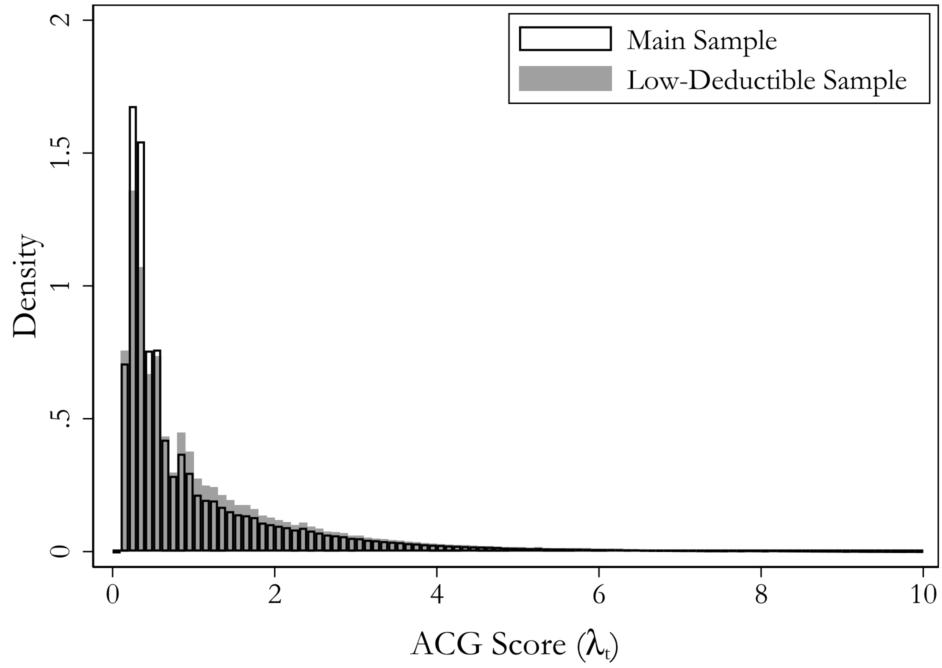


Figure A10: Distribution of λ^* for Main Sample vs. Low-Deductible Plans.

Table A6 shows how clients distribute over different risk categories by age in the low-deductible sample. A comparison with Table 2 confirms that the individuals in the low-deductible sample are in slightly worse health.

Table A6: Health Risk Categories λ by Age Group: Low-Deductible Sample

Age	1 (Healthiest)	2	3	4	5	6	7 (Sickest)
25-30	0.739	0.190	0.049	0.016	0.006	0.001	0.000
30-35	0.672	0.225	0.069	0.025	0.007	0.002	0.000
35-40	0.559	0.282	0.112	0.034	0.011	0.003	0.000
40-45	0.507	0.291	0.141	0.043	0.015	0.003	0.000
45-50	0.406	0.317	0.190	0.060	0.021	0.005	0.001
50-55	0.316	0.311	0.244	0.090	0.030	0.008	0.001
55-60	0.172	0.309	0.320	0.139	0.045	0.013	0.002
60-65	0.093	0.263	0.361	0.190	0.069	0.022	0.003
65-70	0.038	0.200	0.423	0.252	0.072	0.014	0.002
70-75	0.011	0.131	0.403	0.333	0.107	0.015	0.001
75+	0.000	0.055	0.286	0.453	0.179	0.024	0.003

Source: German Claims Panel Data. Sample includes all age groups and uses the ACG© score for the classification.

Table A7 shows the transition probabilities between different health states in the low-deductible sample. The probabilities are very similar to those reported in Table 1.

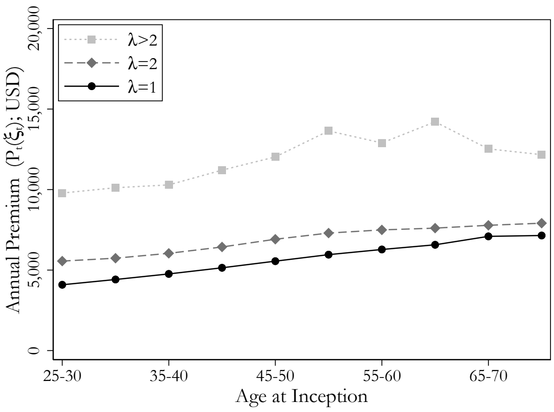
Table A7: Health Risk Category Transitions: Low-Deductible Sample

	λ_{t+1}							
λ_t	1	2	3	4	5	6	7	8 (†)
1	0.802	0.187	0.006	0.002	0.001	0.000	0.000	0.001
2	0.192	0.533	0.232	0.032	0.008	0.001	0.000	0.001
3	0.040	0.168	0.600	0.159	0.026	0.003	0.000	0.003
4	0.017	0.041	0.237	0.553	0.126	0.012	0.000	0.013
5	0.015	0.019	0.034	0.339	0.452	0.102	0.004	0.035
6	0.008	0.013	0.017	0.102	0.313	0.402	0.051	0.094
7	0.000	0.000	0.003	0.027	0.115	0.231	0.426	0.198

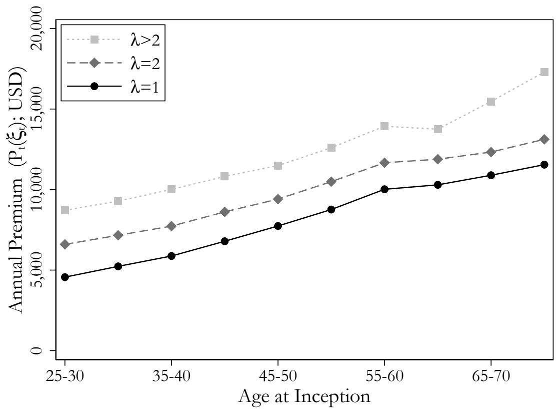
Source: German Claims Panel Data. Sample includes all years, all age groups, and uses the ACG© score for the classification.

A6 Observed vs. Calibrated GLTHI Premium Profiles

Figure A11 compares the (a) calibrated and (b) observed premium profiles for individuals entering their plan at different ages. In both figures, the highest category ($\lambda_t > 2$) is a weighted average calculated according to the actual distribution of λ_t in the different age groups.



(a) Calibrated



(b) Observed

Figure A11: Calibrated vs. Actual Starting Premiums $P_t(\xi_t)$ by Age at Inception

Source: German Claims Panel Data. In Figure A11 (b), the sample includes all years and all health plans, and clients who have been in their contract for 2 to 5 years. We adjusted premiums for the three benefit categories *TOP, PLUS, ECO* and deductible size.

A7 Quantifying Arrow Securities

As discussed in Section 3.3 dynamic long term contracts with one sided commitment can also be implemented by letting the individuals trade state-contingent one-period Arrow securities. Here we provide numerical examples for the GLTHI contract and for the optimal contract.

A7.1 GLTHI Contract

Consider an individual who enters the GLTHI market at age 25 in the healthiest state; $\xi_1 = \Xi_1 = 1$. As shown in Table 3, the GLTHI contract specifies a guaranteed-renewable premium of $P_1(1) = \$3,973$ which represents \$2,499 in excess of expected claims. This front-loading can be reinterpreted as the total amount paid for seven Arrow securities $b_1(k|\Xi_1 = 1)$, $k \in 1, ..., 7$, purchased in period 1 and that would pay 1 dollar if the realized health state in period 2 is k. The prices of each security is given by the corresponding transition probability $\pi(k|1)$. Stacking quantities and probabilities in row vectors; i.e., $\mathbf{b}_1(1) \equiv (b_1(1|\Xi_1 = 1), ..., b_1(7|\Xi_1 = 1)), \boldsymbol{\pi}_1(1) \equiv (\pi_1(1|\Xi_1 = 1), ..., \pi_1(7|\Xi_1 = 1))$, we have:

$$\mathbf{b}_1(1) \times \boldsymbol{\pi}_1(1)' \times \delta + E(m_1|\xi_1 = 1) = P_1(1)$$

Applying formula 3 described in Lemma 1 to our data we get:

$$\mathbf{b_1}(1) = 10^3 \times (1.35, 11.88, 20.62, 28.90, 39.57, 60.25, 202.73)$$

In other words, to replicate the premium path under the GLTHI contract, an individual who starts in the healthiest state purchases a short-term contract in the spot market (at a premium $E(m_1|\xi_1 = 1) = 1,473$), plus a quantity $\mathbf{b}_1$ of seven Arrow securities, at a price $\boldsymbol{\pi}(1) \times \delta$.

Consider now an individual in period 2 with history $\Xi_2 \equiv (\xi_1, \xi_2) = (1, 1)$. As $P_1(1) < P_2(1)$, this individual does not lapse between period 1 and period 2, and continues to pay $\tilde{P}_2([1, 1]) = \$3,973$ in period 2. The individual buys securities in the amount of $\mathbf{b}_2(\xi_2 = 1|\Xi_1 = 1))$ such that

$$\mathbf{b}_2(\xi_2 = 1|\Xi_1 = 1)) \times \boldsymbol{\pi}_2(1)' \times \delta + E(m_2|\xi_2 = 1) - b_1(1|\Xi_1 = 1) = \tilde{P}_2([1, 1]),$$

where $\mathbf{b}_2(\xi_2 = 1|\Xi_1 = 1))$ and $\boldsymbol{\pi}_2(1)$ are analogously defined as $\mathbf{b}_1(1)$ and $\boldsymbol{\pi}_1(1)$, respectively. We get $\mathbf{b}_2(\xi_2 = 1|\Xi_1 = 1)) = 10^3 \times (2.74, 13.25, 22.04, 30.41, 41.22, 62.19, 192.92)$.

Note that the amount of securities bought in any given period t for every future realization of the health status depends on age and on the full history of health statuses up to t.

A7.2 Optimal Contracts

Using the results in A3.1, the vector $b_1^{GHHW}(1)$ shows the quantities of securities purchased for the period-2 contingencies given the individual starts in period 1 that replicate the optimal contract.

$$b_1^{GHHW}(1) = 1000 \times \begin{bmatrix} 0.00 & 3.36 & 8.88 & 15.68 & 25.55 & 44.70 & 181.93 \end{bmatrix}$$

The vector $b_2^{GHHW}([1,1])$ shows the quantities of securities that an individual who starts in period 1 in health state $\xi_1 = 1$, and transitions to state $\xi_2 = 1$ buys for the contingency of transitioning to state $\xi_3 = k$ in period 3.

$$b_2^{GHHW}([1,1]) = 1000 \times \begin{bmatrix} 0.00 & 3.36 & 8.92 & 15.79 & 25.76 & 45.13 & 172.15 \end{bmatrix}$$

A8 Lapsation Rates

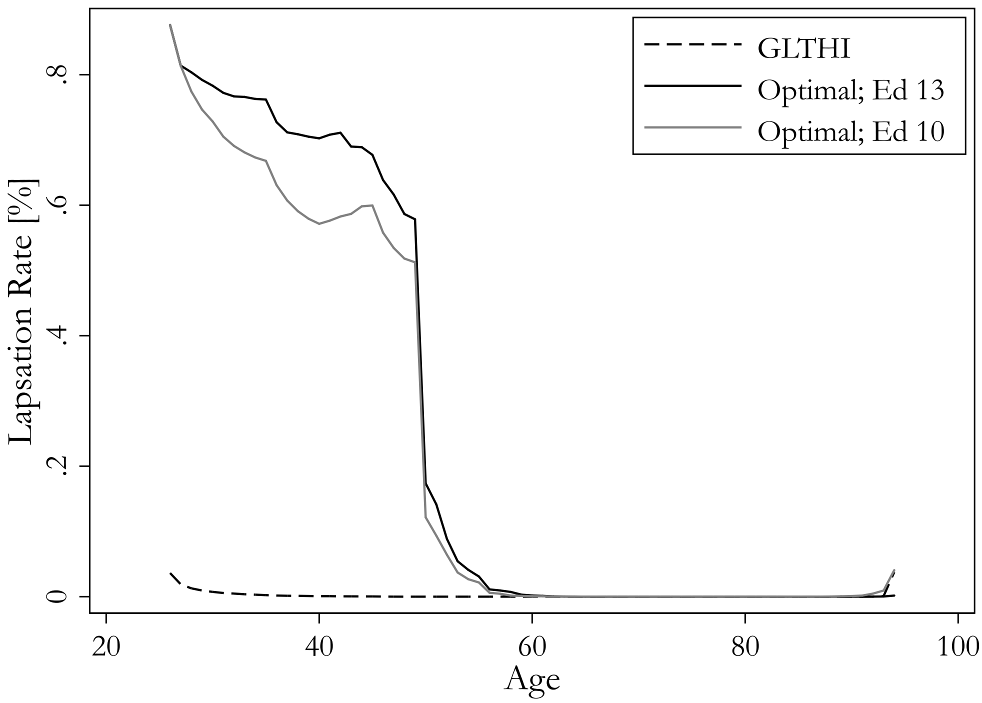


Figure A12: Laspation Rates over the Lifecycle for Different Contracts

Figure A12 compares average lapsation rates under each long-term contract.[A66] As expected, lapsation from GLTHI is extremely low over the entire lifecycle. In contrast, when expected future health improves, the optimal contract results in higher consumption for the healthiest types (and therefore for sicker types too) early in life. Lapsation in the optimal contract decreases substantially in the late 40s. At this point, most individuals have achieved their consumption plateau. Subsequently, consumption remains constant in order to transfer resources intertemporally and to save for old age.

[A66]Lapsing under the optimal contract is defined as an increase in the consumption guarantees. As noted by Ghili et al. (2022), optimal contracts impose a "no-lapsation constraint", so that the consumer will always stay in the same contract. However, an increase in the consumption guarantee specified within a contract can also be interpreted as a lapsation from an equivalent set of guaranteed premium paths. Figure A12 uses this interpretation of lapsing.

A9 Welfare Concepts

We use the concept of lifetime utility U to quantify welfare following, e.g., Ghili et al. (2022):

$$U = \mathbb{E}\left(\sum_{t=t_0}^{T} S_t \delta^{t-t_0} u(c_t)\right)$$

where S_t is an indicator of survival until period t, and c_t is the consumption in period t that is specified by the contract. It may depend on the history of health and income realizations up to t. Expectation is taken over the individual's lifetime health history $(\xi_1, \xi_2, ..., \xi_t)$ and survival.

Certainty Income Equivalent. With a parametric assumption for flow utility $u(.)$, and knowing income y_t, we can summarize welfare with the "*certainty income equivalent*", denoted CE, such that:

$$u(CE) = \frac{\mathbb{E}\left(\sum_{t=t_0}^{T} S_t \delta^{t-t_0} u\left(c_t\right)\right)}{\mathbb{E}\left(\sum_{t=t_0}^{T} S_t \delta^{t-t_0}\right)} \tag{12}$$

This simple expression captures the main trade-offs in health insurance design for lifetime welfare. Lifetime utility is higher when consumption is smoothed across health states and across periods.

First-Best. In particular, the *first-best* consumption level equals the annualized present discounted value of "net income" $y_t - \mathbb{E}(m_t)$, taking into account mortality risk.[A67] This constant optimal consumption level C^* is given by:

$$C^* = \frac{\mathbb{E}\left(\sum_{t=t_0}^{T} S_t \delta^{t-t_0} (y_t - \mathbb{E}(m_t))\right)}{\mathbb{E}\left(\sum_{t=t_0}^{T} S_t \delta^{t-t_0}\right)} \tag{13}$$

Short-Term Contracts. Under a series of actuarially fair *short-term* contracts, the premium in period t with health status ξ_t will simply be $\mathbb{E}(m_t)$. Thus consumption will be $c_t = y_t - \mathbb{E}(m_t|\xi_t)$, and the certainty equivalent CE becomes:

$$u(CE_{ST}) = \frac{\mathbb{E}\left(\sum_{t=t_0}^{T} S_t \delta^{t-t_0} u(y_t - \mathbb{E}(m_t|\xi_t))\right)}{\mathbb{E}\left(\sum_{t=t_0}^{T} S_t \delta^{t-t_0}\right)} \tag{14}$$

[A67] We assume that there is no annuity market, so mortality risk is still considered.

A10 Additional Material for Robustness Analysis

A10.1 Epstein-Zin Recursive Preferences

In the robustness section we evaluate welfare under Epstein-in preferences. Here we provide the derivation for the formula we use to compute the certainty consumption equivalent with Epstein-Zin preferences.

Preferences are defined recursively as

$$V_t = F(c_t, R_t(V_{t+1})),$$

with $R_t(V_{t+1}) = G^{-1}(\mathbb{E}_t G(V_{t+1}))$. As mentioned in the main text, we use the CES aggregator for $F(c,z) = \left((1-\delta)c^{1-1/\psi} + \delta z^{1-1/\psi}\right)^{\frac{1}{1-1/\psi}}$, and incorporate the CARA utility function as $G(c) = u(c) = \frac{1}{\gamma}e^{-\gamma c}$.

Throughout we have assumed that utility is zero if the individual is dead. We can re-interpret V_t as the value of being alive in period t. Under that interpretation, one can write preferences recursively as:

$$V_t = \left((1-\delta)\, c_t^{1-1/\psi} + s_t \delta R_t\left(V_{t+1}\right)^{1-1/\psi}\right)^{\frac{1}{1-1/\psi}} \tag{15}$$

where s_t is the probability of survival between t and $t+1$.

We now derive an expression for the certainty equivalent consumption c for any given value V_t under recursive preferences. Consider the situation in which consumption (while alive) is constant and equal to c. This means that $R_t\left(V_{t+1}\right) = V_{t+1}$, and therefore we can re-write

$$V_t = \left((1-\delta)\, c^{1-1/\psi} + s_t \delta \left(V_{t+1}\right)^{1-1/\psi}\right)^{\frac{1}{1-1/\psi}} \tag{16}$$

Replacing the V_{t+1} in Equation (16) as a function of V_{t+2} yields

$$\begin{aligned} V_t &= \left((1-\delta)\, c^{1-1/\psi} + s_t\delta\left((1-\delta)\, c^{1-1/\psi} + \delta s_{t+1}\left(V_{t+2}\right)^{1-1/\psi}\right)\right)^{\frac{1}{1-1/\psi}} \\ &= \left((1-\delta)\, c^{1-1/\psi} + s_t\delta\,(1-\delta)\, c^{1-1/\psi} + s_t s_{t+1}\delta^2 V_{t+1}^{1-1/\psi}\right)^{\frac{1}{1-1/\psi}} \end{aligned}$$

Iterating forward we can show that

$$\frac{V_t^{1-1/\psi}}{1-\delta} = \sum_{j=t}^{T} c^{1-1/\psi}\delta^{j-t} S_t^j$$

where $S_t^j \equiv \Pi_{k=t}^j s_k$ is the survival probability from t to j. Solving for c, we get an expression defining the certainty equivalent:

$$c = \left(\frac{\frac{V_t^{1-1/\psi}}{1-\delta}}{\sum_{j=t}^T \delta^{j-t} S_t^j} \right)^{\frac{1}{1-1/\psi}} \tag{17}$$

Equation (17) provides the certainty equivalent consumption to a program that provides value V_t.

We are interested in the certainty equivalent taking into account the uncertainty regarding the "birth state" ξ_{t_0}. Denote the value of this lottery V_b. It can be expressed as a function of V_{t_0} (the value at age 25):

$$V_b = G^{-1}(\mathbb{E}_0(G(V_{t_0}(\xi_{t_0})))) \tag{18}$$

where $\mathbb{E}_0()$ takes expectations with respect to the uncertain "birth" state, ξ_{t_0}.

For each contract, we can compute the value $V_{t_0}(\xi_{t_0})$, for each state ξ_{t_0}, *via* backwards induction. Plugging Equation (18) into Equation (17), applied to the initial period t_0 we get the that the consumption certainty equivalent can be expressed as:

$$c = \left(\frac{\frac{\left(G^{-1}(\mathbb{E}_0(G(V_{t_0}(\xi_{t_0}))))\right)^{1-1/\psi}}{1-\delta}}{\sum_{j=t_0}^T \delta^{t-t_0} S_{t_0}^j} \right)^{\frac{1}{1-1/\psi}}, \tag{19}$$

where $\mathbb{E}_0()$ takes expectations with respect to the "birth" state, ξ_{t_0} and S_t^j is the survival probability from t to j. For each contract, we compute $V_{t_0}(\xi_{t_0})$ numerically *via* backwards induction.

A10.2 Initial Health Status Representative of Germany.

In the robustness section we evaluate welfare using a health risk profile that is more representative of the German population as a whole. As mentioned in Appendix A, there is selection into GLTHI contracts for a number of reasons, some imposed by law and some resulting from private incentives of clients and insurance providers. For this reason, it is of great interest to understand if and how our welfare results change when we consider the GKV population.

We start by characterizing and contrasting the two populations, where we draw on two distinct datasets: the representative SOEP panel and a GKV claims dataset. The latter includes a random 5% sample from 2010-11 from one of the largest GKV insurers, 5.6 million enrollees.

Table A8: Socio-Demographics of GKV versus PKV Population

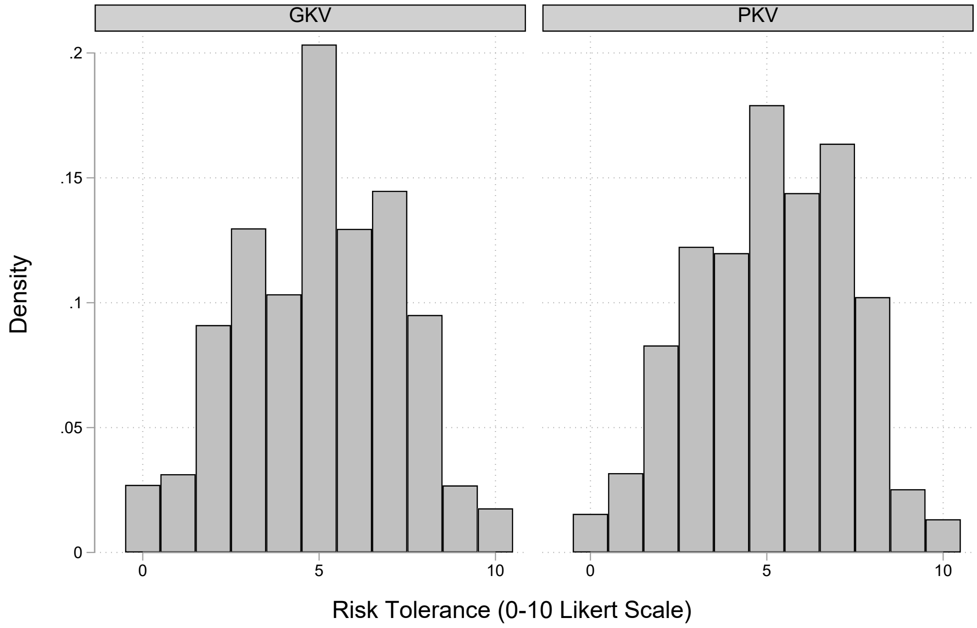
	GKV	PKV		PKV		
			civil servants	white collar	self-employed	non-working
Age	30.110	30.802	30.464	31.540	31.608	29.615
Female	0.467	0.414	0.570	0.195	0.216	0.600
Full-time	0.625	0.817	0.908	0.891	0.931	0.000
Part-time	0.166	0.087	0.092	0.109	0.069	0.000
Dropout	0.017	0.011	0.019	0.000	0.010	0.000
High school	0.315	0.643	0.705	0.723	0.410	0.709
Monthly gross wage	2,611	3,773	3,104	5,285	3,884	2,820
Monthly net wage	1,671	2,581	2,396	3,196	2,466	2,595
Annual household income	24,575	34,858	30,620	46,927	35,540	28,729
Risk tolerance (0-10)	5.040	5.355	4.808	5.885	6.301	4.620
Smoker	0.395	0.292	0.300	0.222	0.385	0.162
BMI	25.032	23.995	23.581	24.536	24.453	23.350
Physical issues, accomplishes less 0.064	0.036	0.036	0.019	0.016	0.090	
Physical limitations	0.053	0.031	0.021	0.012	0.025	0.117
Emotional issues, accomplishes less	0.053	0.030	0.022	0.017	0.012	0.153
Emotional limitations	0.033	0.015	0.012	0.001	0.004	0.085
Hospital stay last year	0.097	0.078	0.073	0.050	0.050	0.235
Hospital nights last year	0.804	0.445	0.374	0.302	0.178	1.613
Doctor visits last quarter	1.780	1.775	2.036	1.247	1.399	2.588
N	23,250	2282	997	465	503	262

Notes: The sample includes 25 to 35 year olds from 2004 to 2016 (except 2005 and 2007). SOEP-provided weights are applied. The number of observations is small for the health behavior, health, and health care utilization measures as they were not surveyed in all years. All income measures have been consistently generated and cleaned by the SOEP team; e.g., *Monthly gross wage* is labeled *labgro* and *Monthly net wage* is labeled *labnet* in SOEP (2018). All income measures are in $2016. When applying t-tests, all means for GKV and PKV are statistically different from one another at conventional levels. Source: SOEP (2018). See Section 4.2 for a detailed discussion of the dataset.

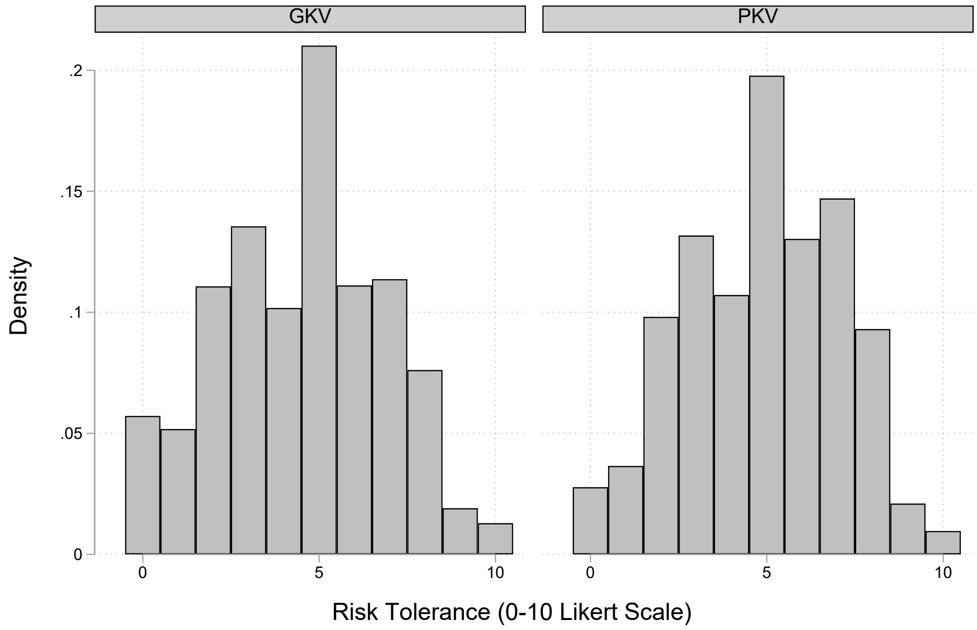
Table A8 provides descriptive statistics from the SOEP, focusing on the 25-35 year-olds. It confirms that the PKV population is positively selected in terms of health and various socioeconomic outcomes. Figure A13 plots the distribution of risk tolerance in the two groups. Even if the risk tolerance is higher among the privately insured on average (5.4 versus 5.0) there is a great deal of overlap of the two distributions, for younger individual as well as for the entire population.

Table A9 compares summary statistics from our PKV claims data to GKV claims data. The GKV sample is seven years older than the PKV sample. The ACG score is also larger on average. When we consider the subsample of 25-35-year-olds in the two rightmost columns, the difference in ACG scores shrinks, but remains larger for the GKV sample. Figure A14 shows the distribution of 2011

ACG scores in the GKV and PKV samples; the left panel plots the raw data for both; the right panel re-weights the GKV observations in order to achieve the same distribution of age by gender as in the PKV dataset.



(a) *25-35 Year Olds*



(b) *All adults*

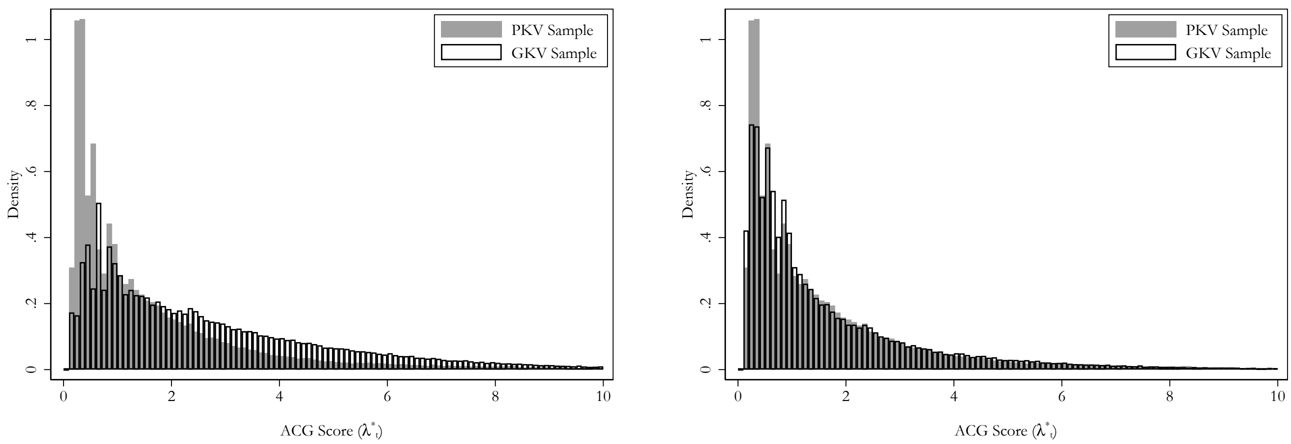
Figure A13: Risk Tolerance: GKV versus PKV population

Notes: The figure shows the distribution of Risk-Tolerance (in a 0-10 Likert Scale) for GKV (public) and PKV (private) enrollees. Panel (a) shows the results for young adults (25-35 year olds) and Panel (b) shows the results for all adults. Source: SOEP (2018).

Table A9: Summary Statistics: PKV versus GKV Claims Data, 2011

	ALL AGED 25+		AGE 25 − 35	
	GKV	**PKV**	**GKV**	**PKV**
Age (in years)	56.3	48.9	29.9	32.2
Female	0.648	0.290	0.558	0.246
ACG Score	2.886	1.745	1.233	0.928
N	226,054	96,511	28,795	9,857

Notes: GKV statistics are based on a 5% random sample of a population of 5.6 million enrollees. Source: German Claims Panel Data and GKV data.



(a) *Unweighted* (b) *Weighted*

Figure A14: Distribution of 2011 ACG Scores: GKV versus PKV population

Note: The figures show histograms of 2011 ACG scores for the GKV and the PKV sample. Panel (a) plots the raw data and panel (b) re-weights the GKV data to get identical distributions of age and gender. Both samples are truncated at an ACG score of 10.

Table A10: Health Risk Categories λ by Age Group: GKV Sample, 2011

Age	1 (Healthiest)	2	3	4	5	6	7 (Sickest)
25-30	0.474	0.316	0.143	0.052	0.012	0.002	0.001
30-35	0.423	0.325	0.161	0.069	0.018	0.004	0.001
35-40	0.296	0.364	0.216	0.088	0.028	0.007	0.001
40-45	0.262	0.366	0.230	0.099	0.033	0.009	0.002
45-50	0.153	0.366	0.290	0.130	0.046	0.013	0.002
50-55	0.127	0.311	0.309	0.165	0.067	0.018	0.003
55-60	0.038	0.254	0.350	0.228	0.097	0.029	0.005
60-65	0.024	0.204	0.333	0.265	0.127	0.042	0.006
65-70	0.025	0.115	0.357	0.330	0.147	0.022	0.003
70-75	0.005	0.069	0.306	0.377	0.205	0.035	0.003
75+	0.000	0.025	0.172	0.407	0.322	0.068	0.005

Notes: The table shows the fraction of individuals in each category by age group. The sample includes all years, all age groups, and uses the ACG scores to construct risk categories λ as explained in the text. Source: Own calculations based on GKV Data.

Table A11 compares our baseline welfare results (Panel (h)) to those of a GKV-only population based on their distribution over starting states (Panel (i)) and to a population approximating the real-world mix of 90% GKV and 10% PKV enrollees (Panel(j)).

Table A11: Welfare under different contracts with initial states representatives of Germany

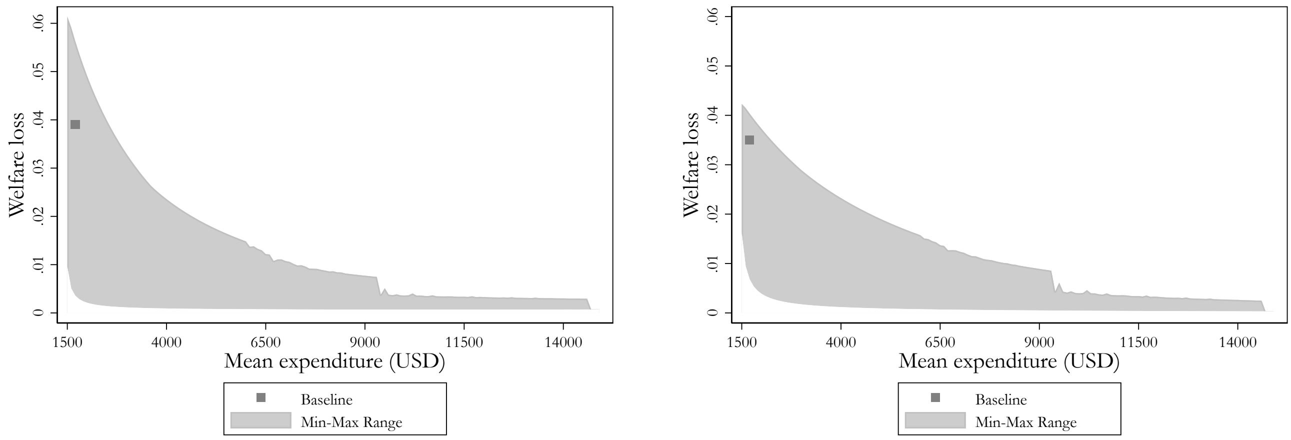
	(1) C^*	(2) C_{ST}	(3) C_{GHHW}	(4) C_{GLTHI}	(5) $\frac{C_{GHHW}-C_{GLTHI}}{C_{GHHW}}$
	Panel (i): $\Delta_0 = \frac{1}{100}[47.47, 31.58, 14.33, 5.16, 1.21, 0.24, 0]$				
Ed 10	22,700	-10,410	19,228	18,966	1.4%
Ed 13	33,878	-2,839	21,633	21,446	0.9%
	Panel (j): $\Delta_0 = \frac{1}{100}[51.63, 29.45, 12.95, 4.65, 1.10, 0.22, 0]$				
Ed 10	22,728	-10,476	19,443	19,155	1.5%
Ed 13	33,906	-2,969	21,885	21,677	0.9%

Notes: The table shows welfare measured by the consumption certainty equivalents in 2016 USD dollars, per capita, per year, separately for two income profiles (see Figure 3). Panel (i) uses the estimated initial probabilities at 25 for the GKV population, and Panel (j) uses a mixture including 90% of individuals in GKV and 10% of individuals in PKV. Columns (1) to (4) show welfare according to the (1) first-best (C^*), (2) a series of short-term contracts (C_{ST}), (3) the optimal contract (C_{GHHW}), and (4) the GLTHI (C_{GHHW}). Column (5)and (6) show how much of the welfare gap between (2) and (1) is closed by the optimal contract and the GLTHI, respectively. Column (7) shows the fraction of the potential benefits from long-term contracting achieved by GLTHI. Column (8) shows the percentage of welfare loss under GLTHI relative to the optimal contract. Source: Own calculations based on German Claims Panel Data, SOEP data.

A10.3 Different Distribution of Starting States

In this section, we asses the robustness of our main results in Table 4 to the initial distribution of starting states. Specifically, we sampled 20 million probability simplices $\tilde{\Delta}_r \in \Delta^7$ from a Dirichlet distribution with concentration parameters equal to the baseline probabilities coming out of the risk classification procedure (including the worst health state 7). The resulting probability simplices span a wide range of expected costs.

Figure A15 plots the resulting welfare difference between GLTHI and the optimal contract in relation to the expected costs associated with each draw. The point "Baseline" corresponds to our baseline estimate in Table 4. According to Figure A15, the welfare loss is bounded from above at about 6% (4% for the less-educated group). The maximum welfare loss decreases in expected expenditures.



(a) *Ed 13* (b) *Ed 10*

Figure A15: Sensitivity Analysis: Distribution of Starting States.

Note: The figures show maxima and minima of GLHTI welfare losses ($\frac{C_{GHHW}-C_{GLTHI}}{C_{GHHW}}$) within increments of $100 of expected expenditure. The underlying distribution is based on 20 million draws from a Dirichlet distribution.

A10.4 Allowing for Savings

In the robustness section we allow for savings under short-term contracts and under the GLTHI contract. Here we provide details of the optimal savings problem.

Individuals solve the following maximization problem:

$$
\begin{aligned}
\max_{c_t} \quad & \mathbb{E}\left(\sum_{t=t_0}^{T} S_t \delta^t u(c_t)\right) \\
\text{s.t.} \quad & a_{t_0} = 0 \\
& a_t \geq 0 \quad \forall t \\
& a_{t+1} = (1+r)a_t + y_t - c_t - \tilde{P}_t(\Xi_t)
\end{aligned}
$$

where $\tilde{P}_t(\Xi_t)$ is the premium paid in period t as a function of an individual's medical history $\boldsymbol{\Xi_t} \equiv (\xi_1, \xi_2, ..., \xi_t)$, and a_t is the level of assets.

Different contracts result in different mappings between an individual's medical history up to period t and an individual's premium in t. Under a series of short-term contracts, only an individual's current health status matters since $\tilde{P}_t(\Xi_t) = \mathbb{E}(m_t|\Xi_t) = \mathbb{E}(m_t|\xi_t)$. In contrast, for a GLTHI contract, the entire medical history matters. Due to guaranteed-renewability, $\tilde{P}_t(\Xi_t)$ is defined recursively: In the first period, $\boldsymbol{\Xi_1} = \xi_1$ and $\tilde{P}_1(\boldsymbol{\Xi_1}) = P_1(\xi_1)$, where Equation (1) in the text defines $P_t(\xi_t)$. In any period $t > 1$, $\tilde{P}_t(\boldsymbol{\Xi_t}) = \min\{\tilde{P}_t(\boldsymbol{\Xi_{t-1}}), P_t(\xi_t)\}$. The state variable in the dynamic program under GLTHI is the guaranteed-renewable premium; its law of motion is given by the probability of qualifying for a lower premium. Note that, in this optimal consumption problem with savings, there is uncertainty regarding net income $y_t - \tilde{P}_t(\Xi_t)$ and mortality risk. Mortality risk implies that individuals may die with positive assets. Therefore, the expected net present value of consumption with optimal savings will be lower than the net present value of resources. Our calculations implicitly assume that individuals do not derive value from bequests.

For a given lifecycle income profile, the dynamic program provides an optimal consumption policy $C_t^*(\xi_t, a_t)$ where a_t is the level of assets carried into period t. The certainty equivalent (CE) of the dynamic problem is equal to:

$$
u(C_{SAV}) = \frac{\mathbb{E}\left(\sum_{t=t_0}^{T} S_t \delta^{t-t_0} u(C_t^*(\xi_t, a_t))\right)}{\mathbb{E}\left(\sum_{t=t_0}^{T} S_t \delta^{t-t_0}\right)} \tag{20}
$$

A10.5 US life-cycle income profiles

To analyze the robustness of our results with respect to using US life-cycle income profile, we employ the representative Panel Study of Income Dynamics (PSID). Figure A16 shows the results of estimating Equation 6 using PSID data for high-school and college-educated. The profiles shown in the Figure are inputs for the welfare calculation using US life-cycle income profiles as described in the main text.

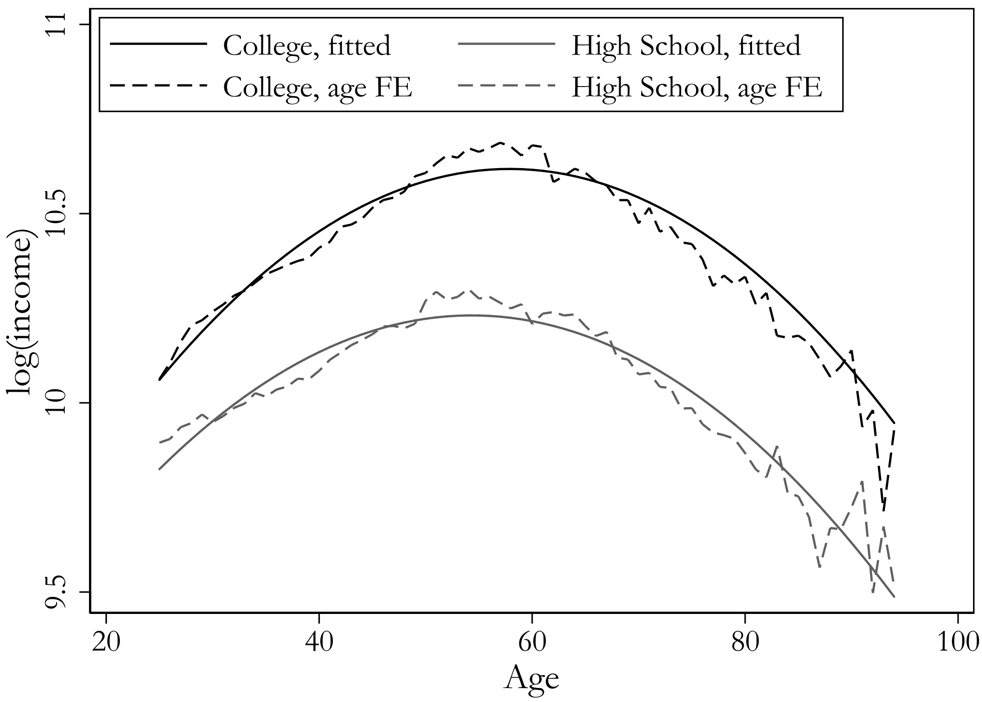


Figure A16: Lifecycle Income Paths for the United States, Nonparametric and Fitted.

Source: Panel Study of Income Dynamics (2018); Frick et al. (2007), years 1984 to 2015. All values in 2016 USD.