Received: 1 April 2021 | Revised: 19 October 2021 | Accepted: 18 February 2022

DOI: 10.1002/hec.4491

Health Economics WILEY

SHORT RESEARCH ARTICLE

Health Economics Letters

When full insurance may not be optimal: The case of restricted substitution

Johannes G. Jaspersen

LMU Munich School of Management, LMU Munich, Ludwig-Maximilians-Universität, Munich, Germany

Correspondence
Johannes G. Jaspersen, LMU Munich School of Management, Ludwig-Maximilians-Universität München, Geschwister-Scholl-Platz 1, Munich 80539, Germany.
Email: jaspersen@lmu.de

Abstract

Even when heavily subsidized, a substantial portion of people choose to forgo purchasing health insurance coverage. In this note, I introduce an explanation for this phenomenon which does not assume choice errors, incorrect beliefs, differently priced uncompensated care, or information asymmetries. When individuals are incapable of freely trading off health and wealth and the initial allocation of goods is suboptimal from their perspective, the standard result of demand for actuarially fair insurance in a single good world does not generalize to the health insurance context. Thus, people might not purchase full health insurance coverage even if it is priced at actuarially fair levels. I argue that this situation is particularly likely to occur in the low-income population, and hence it is relevant for the achievement of universal health coverage.

KEYWORDS

health insurance, underinsurance

JEL CLASSIFICATION

D14, G22, I13

1 | INTRODUCTION

Even at highly subsidized premiums, a notable share of people does not purchase health insurance coverage (Finkelstein et al., 2019). Such behavior is in in contrast with Mossin's (1968) finding that when insurance is available at an actuarially fair rate, expected utility decision-makers should purchase full coverage. While this result was derived for monetary losses, it is commonly applied to health insurance, as well (McGuire, 2011). Because adverse selection alone appears to be insufficient to reconcile empirically documented behavior with canonical theory, explanations typically encompass liquidity constraints (Ericson & Sydnor, 2018), differences in costs of compensated and uncompensated care (Finkelstein et al., 2019), weak administrative capacity (Banerjee et al., 2021), incorrect risk perception (as documented for insurance markets other than health by, e.g., Spinnewijn, 2015) or choice errors (Handel & Kolstad, 2015).

I develop an alternative explanation for the underinsurance phenomenon using a model with expected utility decision-makers who are not subject to any of the above listed confounds. They maximize their utility over health and wealth but are unable to make compensated adjustments to their health status unless they are sick. If decision-makers would like to trade health for wealth in case they are healthy, they will be unwilling to pay for a full recovery in case they are sick. I argue, that such situations may not be uncommon, particularly for the low-income population. If they appear, decision-makers prefer to purchase partial insurance, even if premiums are actuarially fair.

Following Zeckhauser (1970), models of health insurance demand with two attributes have primarily been applied to consumption and health expenditures – a set-up which does not make the trade-off between health and consumption explicit. One exception is the model by Jack and Sheiner (1997). They, however, base their analysis on the indirect utility function and thus implicitly assume that decision-makers can trade-off health and wealth in any desired quantity and direction. Goldman and Philipson (2007) and McGuire (2011) do not make this assumption. While they do not analyze restricted substitution between health and consumption, they show that individuals can prefer partial insurance at actuarially fair prices if the marginal utility of consumption is increasing in health.[1] In a related model, which also requires a positive cross derivative for partial insurance, Cook and Graham (1977) consider insurance for an irreplaceable good – a setting which can easily be applied to health insurance (Zweifel et al., 2009). In contrast to them, I assume that health can be recovered after a shock which makes me focus on a different set of illnesses. Moreover, their model assumes a monetary compensation for a lost good while I assume the more common setting in health insurance of expense-based reimbursement.

Restricted substitution removes the requirement for a positive cross derivative of the utility function. Rather, my results hold for a large set of possible utility functions with only weak conditions on the cross derivative. This is important, because empirical evidence does not suggest a universally positive cross-derivative. While Viscusi and Evans (1990), Finkelstein et al. (2013) and Blundell et al. (2020) estimate a positive effect of health on marginal utility, Evans and Viscusi (1991) find no such effect for transitory changes in health. Likewise, Viscusi (2019) argues for a positive cross derivative, but only if the health shock is severe.

2 | RESULTS

Decision-makers maximize utility over consumption and health: $U(c, h)$. Using the notation $\partial U/\partial c = U_c$, $\partial^2 U/\partial c^2 = U_{cc}$, $\partial^3 U/\partial c^3 = U_{ccc}$, $\partial U/\partial h = U_h$, $\partial^2 U/\partial h^2 = U_{hh}$ and $\partial^2 U/\partial c \partial h = U_{ch} = U_{hc}$, I assume positive and decreasing marginal utility of both health and consumption as well as prudence over money (U_c, $U_h > 0$,U_{cc}, $U_{hh} < 0$, and $U_{ccc} > 0$).[2] Further, I assume that the cross derivative of the utility function does not overwhelm the effect of marginal utility on preferences when adjusting for relative prices. As is detailed below, decision-makers can recover from a bad health event at monetary costs q. Formally, the assumption on the cross derivative implies that $U_{ch} > qU_{cc}$ and $qU_{ch} > U_{hh}$. As is explained in detail by Chung et al. (2019), this assumption is without loss of much generality and should cover the vast majority of decision-makers in the health insurance market. As more of a technical assumption, I further assume that the cross derivative is not too large in magnitude such that insurance demand and out-of-pocket payments are determined in a concave decision problem.

Decision-makers have initial wealth w_0 and health h_0 and face the risk that their health decreases to $h_0 - L$ with probability p. When sick, decision-makers can purchase healthcare for a price q. For ease of exposition, I focus on the case in which health gains are linear in health care expenses. However, the intuition of the analysis carries over to non-linear health production functions as I show in Online Appendix C. In case the insured are healthy, they cannot change the allocation of consumption and health. That is, the substitution between the two arguments of the utility function is restricted to one direction. Also, decision-makers cannot increase the health status above the initial level of h_0. Insurance against the incursion of medical costs is available at an actuarially fair price. Decision-makers can choose the degree of insurance coverage, denoted α, on a continuous scale between 0 and 1. The insurance contract has an expense-based reimbursement system such that decision-makers have to spent the indemnity on healthcare. The actuarially fair premium, payable in both states of the world, is αpqL.

Decision-makers choose the amount of health insurance to buy. In case they are sick and have purchased less than full insurance, they can also purchase health care out-of-pocket, denoted t (such that $0 \leq t \leq 1 - \alpha$). The objective function reads

$$\max_{\alpha \geq 0, t \geq 0} \{ EU(\alpha, t) = (1-p)U(w_0 - \alpha pqL, h_0) + pU(w_0 - \alpha pqL - tqL, h_0 - (1 - \alpha - t)L) \}.$$

Proposition 1 states the main result.

PROPOSITION 1 *If $qU_c(w_0, h_0) > U_h(w_0, h_0)$ and insurance is actuarially fair,*

(i) decision-makers do not purchase full insurance, and
(ii) if $qU_c(w_0, h_0 - L) < U_h(w_0, h_0 - L)$, $t > 0$ if and only if $U_{ch} > 0$.

All proofs are provided in the appendix. $qU_c(w_0, h_0) > U_h(w_0, h_0)$ implies that in their initial allocation (w_0, h_0), decision-makers would be willing to decrease their health status if they could gain the equivalent treatment costs in additional consumption.[3] The intuition of item (i) is thus that decision-makers do not want to purchase full insurance, because it forces them to allocate more money into the treatment cost than they want to spend on health. Item (ii) follows, because if consumption is worth less in bad health, decision-makers allocate more healthcare costs to that state of the world. The intuition behind this result is equal to that behind the result of Cook and Graham (1977) and Zweifel et al. (2009). The result reflects the fact that out-of-pocket payments are not uncommon, particularly in the low-income population. $qU_c(w_0, h_0 - L) < U_h(w_0, h_0 - L)$ simply establishes that some amount of treatment is demanded in case of sickness.

Under which circumstances will the condition $qU_c(w_0, h_0) > U_h(w_0, h_0)$ most likely hold? Healthcare is generally considered a normal good (see, e.g., Evans & Viscusi, 1993; Alfonso et al., 2016). Thus, the condition is most likely to be true in the low-income population. For the given model, I derive this result formally in the next proposition which shows that all other things remaining equal, $qU_c(w_0, h_0) > U_h(w_0, h_0)$ is more likely to be true as w_0 decreases.

PROPOSITION 2 *Ceteris paribus, the expression $qU_c(w_0, h_0) - U_h(w_0, h_0)$ is decreasing in w_0.*

As individuals obtain more wealth, the marginal utility of non-medical consumption decreases. The cost of improving one's health is unaffected by wealth. Because the marginal utility of health at the very least decreases less than the marginal utility of consumption, spending money on healthcare becomes more attractive. With decreasing wealth, $qU_c(w_0, h_0) - U_h(w_0, h_0)$ is increasing and $qU_c(w_0, h_0) > U_h(w_0, h_0)$ is more likely to hold. Hence, underinsurance due to restricted substitution between health and wealth is more likely in the low-income population.

Figure 1 gives a numerical illustration of the results. I set $h_0 = q = 1$, $L = 0.8$, $p = 0.1$ and $U(c, h) = c^\gamma + h^\gamma + kc^\gamma h^\gamma$, such that U_{ch} has the same sign as k. To ensure risk aversion for both health and wealth, I set $\gamma = 0.7$.[4] The figure shows the optimal insurance coverage and out-of-pocket payments for wealth ranging between 0 and $1.5qh_0$. The sufficient condition for Proposition 1 is fulfilled for any w_0 smaller $1qh_0$.

Irrespective of U_{ch}, decision-makers purchase full insurance if initial wealth is somewhat larger than $1qh_0$.[5] Below this point, coverage drops below 1 and is decreasing in wealth. This demonstrates that the effect described in Proposition 1 is not minor and may even lead decision-makers to completely forgo purchasing health insurance. Also, the proposed mechanism increases in relevance as the decision-makers' wealth decreases.

3 | DISCUSSION

Both Proposition 2 and Figure 1 imply that restricted substitution will most likely influence how the low-income population purchases health insurance. This corresponds to empirical findings, which document lacking health insurance demand for the low-income population (Finkelstein et al., 2019), while finding no such results for higher income strata

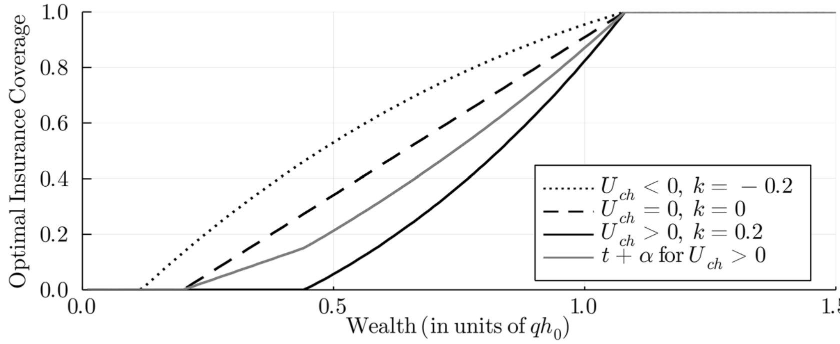

FIGURE 1 Optimal levels of insurance demand (α) and out-of pocket payments (t) in a setting where $U(c, h) = c^\gamma + h^\gamma + kc^\gamma h^\gamma$, $h_0 = q = 1$, $p = 0.1$, $L = 0.8$ and $\gamma = 0.7$

(Hackmann et al., 2015). Take-up of subsidized health insurance might thus be increased if the subsidies were accompanied by a social safety net which protects individuals from situations with very low levels of wealth.

My results depend on the preference condition $qU_c(w_0, h_0) < U_h(w_0, h_0)$, direct empirical evidence for which is rare. Sabat and Gallagher (2019) show that the low-income population often foregoes purchasing healthcare even in the presence of liquidity reserves. In their data, foregoing healthcare also is, on average, the first drastic savings measure in financial hardship.[6] Gross et al. (2021) find that individuals may not fill drug prescriptions when they are low on liquidity even if the copayments for these prescriptions are so low that liquidity constrains are unlikely. Income is also a decisive factor in the monetary evaluation of quality-adjusted life years (Bobinac et al., 2010) and the take-up of dental care (Allin et al., 2009). Both results give credibility to the preference condition being more likely for the low-income population. More indirectly, the fact that some people are willing to trade-off their (expected) health for wealth is also evident by the willingness to accept potentially health-threatening jobs. Viscusi and Aldy (2003) show that such behavior is more common in the low-income population. Lastly, Di Matteo (2003) shows that the income elasticity of health expenditures in the US low-income population is above 1. This suggests that willingness to pay for health care drops faster than income at low income brackets, making $qU_c(w_0, h_0) > U_h(w_0, h_0)$ more likely at low levels of w_0.

The results of this note rely on the assumptions that $U_{ccc} > 0$ and that the magnitude of U_{ch} is not too large. These two assumptions are sufficient conditions for a fulfilled second order condition and thus a concave decision problem, but they are not necessary. Moreover, the goal of this note is not to provide a universal result on health insurance behavior, but rather to offer a potential explanation for an observed result under a reasonable set of assumptions. It is unnecessary that the assumptions hold for all decision-makers – they merely need to hold for a subset of the population.

While the above model uses a simple set-up for ease of exposition, two extensions are possible. The first covers the case in which health can only be restored up to some level $\bar{h} \leq h_0$. This requires adjustment of the preference condition such that it holds in the best possible health state which can still be achieved when sick: $qU_c(w_0, \bar{h}) > U_h(w_0, \bar{h})$. The second extension, detailed in Online Appendix C, covers a non-linear health production function.

ACKNOWLEDGEMENT

I would like to thank the editor Eleonora Fichera as well as two anonymous reviewers for their feedback and Benjamin Collier, Theresa Eden, Richard Peter, Marc Ragin and Casey Rothschild for valuable comments. Joëlle Näger provided valuable research assistance. All remaining errors are mine.

CONFLICT OF INTEREST

The author reports no conflicts of interest (financial or otherwise) and no competing interests.

DATA AVAILABILITY STATEMENT

The code to create the numerical data that support the findings of this study are openly available on the author's homepage at https://sites.google.com/jaspersen.com/jgj/home.

ORCID

Johannes G. Jaspersen https://orcid.org/0000-0002-3599-8988

ENDNOTES

1 It is worth mentioning that this idea is not explored in detail in either Goldman and Philipson (2007) or McGuire (2011). The former focus on the implications of moral hazard, while the latter study focuses on the case in which marginal utility of consumption is decreasing in wealth.

2 Risk aversion over health and consumption are standard assumptions and prudence has been well established for monetary stakes (e.g., Ebert & Wiesen, 2011).

3 This can also be expressed as $U_c(w_0, h_0)/U_h(w_0, h_0) > 1/q$ to clarify that it is a condition on the marginal rate of substitution between health and consumption.

[4] Results for alternative values of γ, p and L can be found in Online Appendix B. None of the parameter choices materially affect the results.

[5] Optimal coverage is partial for some wealth levels above qh_0 because $qU_c(w_0, h_0) > U_h(w_0, h_0)$ is a sufficient but not necessary condition. The necessary and sufficient condition is $qU_c(w_0 - pqL, h_0) > U_h(w_0 - pqL, h_0)$.

[6] Other analyzed measures are skipping meals, skipping rent payments and not paying utility bills.

REFERENCES

Alfonso, Y. N., Ding, G., & Bishai, D. (2016). Income elasticity of vaccines spending versus general healthcare spending. *Health Economics*, *25*(7), 860–872.

Allin, S., Masseria, C., & Mossialos, E. (2009). Measuring socioeconomic differences in use of health care services by wealth versus by income. *American Journal of Public Health*, *99*(10), 1849–1855.

Banerjee, A., Finkelstein, A., Hanna, R., Olken, B. A., Ornaghi, A., & Sumarto, S. (2021). The challenges of universal health insurance in developing countries: Experimental evidence from Indonesia's national health insurance. *American Economic Review*, *111*(9), 3035–3063.

Blundell, R., Borella, M., Commault, J., & De Nardi, M. (2020). *Why does consumption fluctuate in old age and how should the government insure it?* Working Paper No. 27348. National Bureau of Economic Research.

Bobinac, A., Van Exel, N. J. A., Rutten, F. F., & Brouwer, W. B. (2010). Willingness to pay for a quality-adjusted life-year: The individual perspective. *Value in Health*, *13*(8), 1046–1055.

Chung, H.-K., Glimcher, P., & Agniezka, T. (2019). An experimental comparison of risky and riskless choice – limitations of prospect theory and expected utility theory. *American Economic Journal: Microeconomics*, *11*(3), 34–67.

Cook, P. J., & Graham, D. A. (1977). The demand for insurance and protection: The case of irreplaceable commodities. *Quarterly Journal of Economics*, *91*(1), 143–156.

Di Matteo, L. (2003). The income elasticity of health care spending. *The European Journal of Health Economics*, *4*(1), 20–29.

Ebert, S., & Wiesen, D. (2011). Testing for prudence and skewness seeking. *Management Science*, *57*(7), 1334–1349.

Ericson, K. M., & Sydnor, J. R. (2018). Liquidity constraints and the value of insurance. *Working paper*. National Bureau of Economic Research, No. 24993.

Evans, W. N., & Viscusi, W. K. (1991). Estimation of state-dependent utility functions using survey data. *The Review of Economics and Statistics*, *73*(1), 94–104.

Evans, W. N., & Viscusi, W. K. (1993). Income effects and the value of health. *Journal of Human Resources*, *28*(3), 497–518.

Finkelstein, A., Hendren, N., & Shepard, M. (2019). Subsidizing health insurance for low-income adults: Evidence from Massachusetts. *American Economic Review*, *109*(4), 1530–1567.

Finkelstein, A., Luttmer, E. F., & Notowidigdo, M. J. (2013). What good is wealth without health? The effect of health on the marginal utility of consumption. *Journal of the European Economic Association*, *11*(1), 221–258.

Goldman, D., & Philipson, T. J. (2007). Integrated insurance design in the presence of multiple medical technologies. *American Economic Review*, *97*(2), 427–432.

Gross, T., Layton, T. J., & Prinz, D. (2021). *The liquidity sensitivity of healthcare consumption: Evidence from social security payments*. American Economic Review: Insights, in press.

Hackmann, M. B., Kolstad, J. T., & Kowalski, A. E. (2015). Adverse selection and an individual mandate: When theory meets practice. *American Economic Review*, *105*(3), 1030–1066.

Handel, B. R., & Kolstad, J. T. (2015). Health insurance for" humans": Information frictions, plan choice, and consumer welfare. *American Economic Review*, *105*(8), 2449–2500.

Jack, W., & Sheiner, L. (1997). Welfare-improving health expenditure subsidies. *American Economic Review*, *87*(1), 206–221.

McGuire, T. G. (2011). Demand for health insurance. In *Handbook of Health Economics* (Vol. 2, pp. 317–396). Elsevier.

Mossin, J. (1968). Aspects of rational insurance purchasing. *Journal of Political Economy*, *76*(4, Part 1), 553–568.

Sabat, J., & Gallagher, E. (2019). *Rules of thumb in household savings decisions: Estimation using threshold regression*. Universidad Diego Portales.

Spinnewijn, J. (2015). Unemployed but optimistic: Optimal insurance design with biased beliefs. *Journal of the European Economic Association*, *13*(1), 130–167.

Viscusi, W. K. (2019). Utility functions for mild and severe health risks. *Journal of Risk and Uncertainty*, *58*(2), 143–166.

Viscusi, W. K., & Aldy, J. E. (2003). The value of a statistical life: A critical review of market estimates throughout the world. *Journal of Risk and Uncertainty*, *27*(1), 5–76.

Viscusi, W. K., & Evans, W. N. (1990). Utility functions that depend on health status: Estimates and economic implications. *The American Economic Review*, *80*(9), 353–374.

Zeckhauser, R. (1970). Medical insurance: A case study of the tradeoff between risk spreading and appropriate incentives. *Journal of Economic Theory*, *2*(1), 10–26.

Zweifel, P., Breyer, F., & Kifmann, M. (2009). *Health economics* (2nd ed.). Springer.

How to cite this article: Jaspersen, J. G. (2022). When full insurance may not be optimal: The case of restricted substitution. *Health Economics*, *31*(6), 1249–1257. https://doi.org/10.1002/hec.4491

APPENDIX A: PROOFS OF PROPOSITIONS

A.1 | Proof of Proposition 1

I will abbreviate $w_0 - \alpha pqL, h_0$ as y_I and $w_0 - \alpha pqL - tqL, h_0 - (1-\alpha-t)L$ as y_{II} such that y_I and y_{II} denote wealth and health in the state without and with a health shock, respectively.

Item (i): The decision-maker maximizes expected utility over $\alpha \geq 0$ and $t \geq 0$ with the restrictions $\alpha + t \leq 1$ such that the set of possible choices is closed, convex and a lattice. Further, $\partial^2 EU/\partial\alpha\partial - t = pL^2[q(1+p)U_{ch}(y_{II}) - U_{hh}(y_{II}) - pq^2U_{cc}(y_{II})] = pL^2[pq(U_{ch}(y_{II}) - qU_{cc}(y_{II})) + (qU_{ch}(y_{II}) - U_{hh}(y_{II}))] > 0$. Such that expected utility has increasing differences in $(\alpha, -t)$.

If we set $t = 0$, we can solve for an optimal degree of coinsurance without out-of-pocket payments, denoted $\hat{a}$. It fulfills $\partial EU/\partial\alpha|_{t=0} = U_h(w_0 - \hat{a}pqL, h_0 - (1-\hat{a})L) - q[(1-p)U_c(w_0 - \hat{a}pqL, h_0) + pU_c(w_0 - \hat{a}pqL, h_0 - (1-\hat{a})L)] = 0$. We establish that $\partial^2 EU/\partial\alpha^2|_{t=0} < 0$, by seeing that $q^2p[(1-p)U_{cc}(y_I) + pU_{cc}(y_{II})] - 2pqU_{ch}(y_{II}) + U_{hh}(y_{II}) < q^2pU_{cc}(y_{II}) - 2pqU_{ch}(y_{II}) + U_{hh}(y_{II}) = qp[qU_{cc}(y_{II}) - U_{ch}(y_{II})] + [U_{hh} - qpU_{ch}] < 0$, where the first inequality follows from $U_{ccc} > 0$ and the second one follows from our assumptions on U_{ch}. $\partial^2 EU/\partial\alpha^2|_{t=0} < 0$ shows that the second order condition is fulfilled and thus $\hat{a} < 1$ if $\partial EU/\partial\alpha|_{t=0,\alpha=1} < 0$. $\partial EU/\partial\alpha|_{t=0,\alpha=1} = U_h(w_0 - pqL, h_0) - qU_c(w_0 - pqL, h_0)$. From $qU_c(w_0, h_0) > U_h(w_0, h_0)$, we know that $U_h(w_0, h_0) - qU_c(w_0, h_0) < 0$. Further, due to $U_{ch} > qU_{cc}$, $U_h(w_0, h_0) - U_h(w_0 - qpL, h_0) > qU_c(w_0, h_0) - qU_c(w_0 - pqL, h_0)$. Rearranging renders $\partial EU/\partial\alpha|_{\alpha=1} < U_h(w_0, h_0) - qU_c(w_0, h_0) < 0$ and thus $\hat{a} < 1$.

Because EU has increasing differences in $(\alpha, -t)$, α will be lower with increasing t. We thus know that the optimal level of α when t is not restricted to 0 is weakly smaller than the optimal level of α when t is restricted to 0. It follows that $\alpha^* \leq \hat{a} < 1$.

Item (ii): We now consider the decision-problem using a Lagrangian approach with two inequality constraints. We first note that EU is concave in (α, t) by seeing that due to $U_{ccc} > 0$ it holds that $\partial^2 EU/\partial\alpha^2 = pL^2[q^2pEU_{cc} - 2pqU_{ch}(y_{II}) + U_{hh}(y_{II})] < pL^2[qp(qU_{cc}(y_{II}) - U_{ch}(y_{II})) + (U_{hh}(y_{II}) - pqU_{ch}(y_{II}))] < 0$, that $\partial^2 EU/\partial t^2 = pL^2[q(qU_{cc}(y_{II}) - U_{ch}(y_{II}) + (U_{hh}(y_{II}) - qU_{ch}(y_{II}))] < 0$ and that the determinant of the Jacobian is positive. The latter can be seen by $D = \partial^2 EU/\partial\alpha^2 \times \partial^2 EU/\partial t - (\partial^2 EU/\partial\alpha\partial t)^2 = p^2L^4[pq^2EU_{cc} - 2pqU_{ch}(y_{II}) + U_{hh}(y_{II})][q^2U_{cc}(y_{II}) - 2qU_{ch}(y_{II}) + U_{hh}(y_{II})] - p^2L^4[pq^2U_{cc}(y_{II}) - q(1+p)U_{ch}(y_{II}) + U_{hh}(y_{II})]^2$. We show that $D/p^2L^4 > 0$ and thus $D > 0$. Due to $U_{ccc} > 0$, $D/p^2L^4 > [pq^2U_{cc}(y_{II}) - 2pqU_{ch}(y_{II}) + U_{hh}(y_{II})][q^2U_{cc}(y_{II}) - 2qU_{ch}(y_{II}) + U_{hh}(y_{II})] - [pq^2U_{cc}(y_{II}) - q(1+p)U_{ch}(y_{II}) + U_{hh}(y_{II})]^2 = pq^2U_{cc}[(1-p)q^2U_{cc} - 2pqU_{ch}] + U_{hh}[(1-p)q^2U_{cc} - 2pqU_{ch}] - q^2(1-p)^2(U_{ch})^2$ which is positive if U_{ch} is not too large in absolute magnitude, as is assumed.

Concavity of EU in (α, t) and maximization over a closed and convex set ensure that the Karush-Kuhn-Tucker (KKT) conditions of the Lagrangian are both necessary and sufficient for a maximization. These conditions are given by

(1): $\partial\mathcal{L}/\partial\alpha = pL[U_h(y_{II}) - q(1-p)U_c(y_I) - qpU_c(y_{II})] + \lambda_1 = 0$,
(2): $\partial\mathcal{L}/\partial t = pL[U_h(y_{II}) - qU_c(y_{II})] + \lambda_2 = 0$,
(3)–(6): $\lambda_1, \lambda_2 \geq 0, \lambda_1\alpha = 0, \lambda_2 t^* = 0$.

We first establish that under the conditions of the proposition, α^* and t^* cannot be 0 at the same time. This would imply that $U_h(y_{II}) - qU_c(y_{II}) = U_h(w_0, h_0 - L) - qU_c(w_0, h_0 - L) > 0$, such that KKT conditions (2) and (4) contradict one another. Thus, either α^* or t^* or both need to be positive.

We now show that $U_{ch} \leq 0$ and $t > 0$ cannot lead to fulfilled KKT conditions. From $t > 0$, $\lambda_2 = 0$ follows. This implies $pLU_h(y_{II}) = pLqU_c(y_{II})$ from (2) which we substitute into (1) to obtain $pLq(1-p)[U_c(y_{II}) - U_c(y_I)] + \lambda_1 = 0$. (3) thus implies that $U_c(y_{II}) - U_c(y_I) \leq 0$. However, we know from item (i) that $\alpha < 1$, making consumption in y_{II} smaller than in y_I. We also know that because health in y_{II} cannot be larger than in y_I. Combining both observations, $U_{cc} < 0$, and $U_{ch} > 0$ shows that $U_c(y_{II}) > U_c(y_I)$ and thus renders a contradiction.

Lastly, we show that $U_{ch} > 0$ and $t^* = 0$ cannot fulfill the KKT conditions, but that they can be fulfilled for $t^* > 0$. Starting at $t^* = 0$, we know $\alpha^* > 0$ and thus $\lambda_1 = 0$. From (1), we can then derive $U_h(y_{II}) = q(1-p)U_c(y_I) - qpU_c(y_{II})$. Substituting into (2) renders $pLq(1-p)[U_c(y_I) - U_c(y_{II})] + \lambda_2 = 0$. However, at $t^* = 0$ this implies $U_c(w_0 - \alpha^*pqL, h_0) - U_c(w_0 - \alpha^*pqL, h_0 - (1-\alpha^*)L) \leq 0$. This is impossible for $U_{ch} > 0$ as long as $\alpha^* < 1$, which we know is true from item (i). Thus $U_{ch} > 0$ contradicts $t^* = 0$. For t^*v > 0, $\lambda_2 = 0$ and thus $U_h(y_{II}) = qU_c(y_{II})$. Substituting into (1) implies $U_c(y_{II}) - U_c(y_I) = U_c(w_0 - \alpha^*pqL - t^*qL, h_0 - (1-\alpha^* - t^*q)L) - U_c(w_0 - \alpha^*pqL, h_0) \leq 0$. At $t^* = 0$, U_{ch} implies $U_c(y_{II}) - U_c(y_I) < 0$ and because $U_c(y_{II}) - U_c(y_I)$ is decreasing in t, this will also be true for $t^* > 0$.

A.2 | Proof of Proposition 2

The derivative of the expression toward w_0 is $qU_{cc}(w_0, h_0) - U_{ch}(w_0, h_0)$. This is negative under the assumptions made on the preferences.

APPENDIX B: NUMERICAL RESULTS WITH ALTERNATIVE PARAMETRIZATIONS

In this online appendix, I show the results of the numerical illustration for different values of γ, p and L. Figure B1 shows the results for $\gamma = 0.2$ and $\gamma = 0.5$. Lower values of γ imply higher levels of utility curvature for consumption and health. Results are very similar to those in the main analysis. Decreasing γ leads to less distinguished results for the different values of k. Additionally, smaller levels of γ decrease the share of out-of-pocket payments relative to that of expenses covered by insurance.

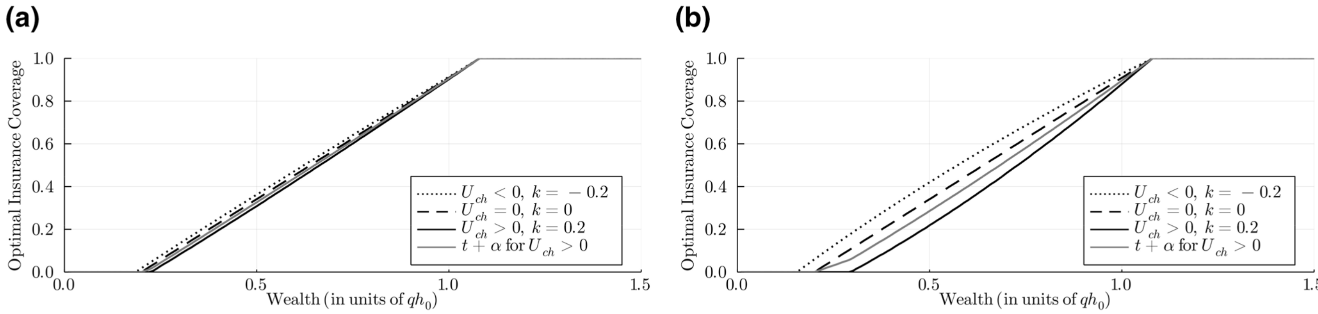

FIGURE B1 The graph shows the results of Figure 1 for different values of γ. Lines show the numerically determined optimal levels of insurance demand (α) and out-of pocket payments (t) in a setting where $U(c,h) = c^{\gamma} + h^{\gamma} + kc^{\gamma}h^{\gamma}$, $h_0 = q = 1$, $p = 0.1$ and $L = 0.8$. (a) Results for $\gamma = 0.2$. (b) Results for $\gamma = 0.5$

Figures B2–B4 demonstrate the results for different levels of p and L. In Figure B2, I keep the level of the loss constant at 0.8, while increasing p to 0.2 and decreasing it to 0.05. In Figure B3, the probability is kept constant, while the loss is increased to 1.0 (the maximal possible level) and decreased to 0.4. Lastly, Figure B4 keeps the expected value of the health loss constant at 0.08 as in the main analysis, but increases the probability of loss, while simultaneously decreasing the loss amount. Across all three figures, the probability of loss seems to have very little influence on the results. In contrast, lowering the loss amount makes the results more extreme in the sense that insurance demand drops more quickly as wealth decreases.

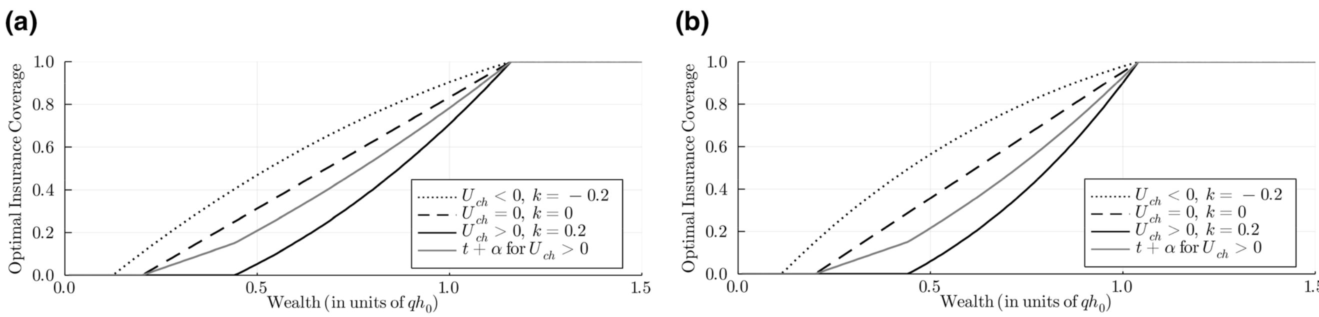

FIGURE B2 The graph shows the results of Figure 1 for different values of p. Lines show the numerically determined optimal levels of insurance demand (α) and out-of pocket payments (t) in a setting where $U(c,h) = c^{\gamma} + h^{\gamma} + kc^{\gamma}h^{\gamma}$, $h_0 = q = 1$, $\gamma = 0.7$ and $L = 0.8$. (a) Results for $p = 0.2$ and $L = 0.8$. (b) Results for $p = 0.05$ and $L = 0.8$

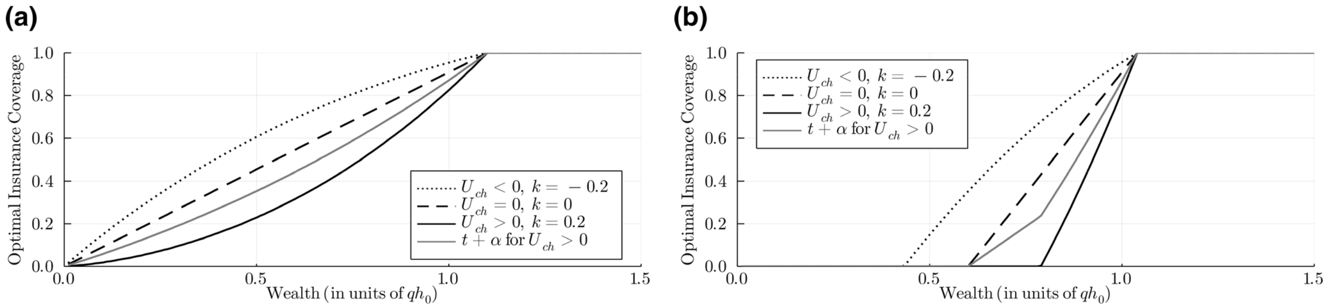

FIGURE B3 The graph shows the results of Figure 1 for different values of L. Lines show the numerically determined optimal levels of insurance demand (α) and out-of pocket payments (t) in a setting where $U(\text{c,h}) = c^\gamma + h^\gamma + kc^\gamma h^\gamma$, $h_0 = q = 1$, $\gamma = 0.7$ and $p = 0.1$. (a) Results for $p = 0.1$ and $L = 1.0$ (b) Results for $p = 0.1$ and $L = 0.4$

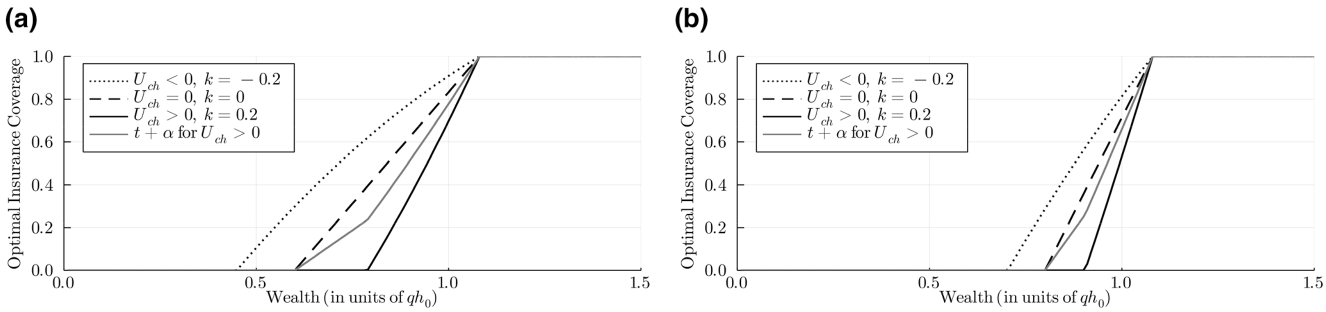

FIGURE B4 The graph shows the results of Figure 1 for different values of p and L but the same expected value of the loss as in the main analysis. Lines show the numerically determined optimal levels of insurance demand (α) and out-of pocket payments (t) in a setting where $U(\text{c,h}) = c^\gamma + h^\gamma + kc^\gamma h^\gamma$, $h_0 = q = 1$, $\gamma = 0.7$ and $pL = 0.08$. (a) Results for $p = 0.2$ and $L = 0.4$. (b) Results for $p = 0.4$ and $L = 0.2$

APPENDIX C: EXTENSION TO NON-LINEAR HEALTH PRODUCTION FUNCTION

A potential criticism of the model analyzed in the main text is that a linear benefit function of monetary payments for health care is unrealistic. It is indeed the case that cheaper procedures or medications can often achieve some amount of health gains but fall short of achieving health levels close to the optimum, which require more expensive treatments. An arthritic joint in an elderly patient, for example, can be treated with pain medication or with surgical insertion of a prothesis. While both restore the mobility of the patient, the cheaper treatment with pure medication has side effects on liver and kidneys, while the operation is disproportionally more expensive, but has fewer side effects. For medication, diabetes treatments pose a good example of a concave benefit function. Diabetes can be treated with human insulin or human-analog insulin. The cheaper human insulin can be used for medication, but works in a delayed fashion, allowing less precise steering of the insulin level, which requires a lower target level of insulin than is used when medicating with the faster acting and more easily steerable human-analog insulin. Thus, some increases from a bad health state can be achieved by the cheap medication, but improvements beyond these initial increases are disproportionally more expensive.

I model such a health production function formally, by assuming that monetary payments transfer into the health status concavely with function r: $[0,1] \rightarrow [0,1]$ with $r(0) = 0$, $r(1) = 1$, $r' > 0$ and $r'' \leq 0$. This also changes the formal form of the condition that the exchange rate adjusted cross derivative of the utility function does not overwhelm the effect of marginal utility on preferences. Specifically, this condition now dictates that $q/r'(x)U_{cc}(c, h_0 - (1 - r(x))L) < U_{ch}(c, h_0 - (1 - r(x))L)$ and $U_{hh}(c, h_0 - (1 - r(x))L) < q/r'(x)U_{ch}(c, h_0 - (1 - r(x))L)$ for all levels of c and x.

The objective function of this extended model is

$$\max_{\alpha \geq 0, t \geq 0} \{EU(\alpha) = (1-p)U(w_0 - \alpha pqL, h_0) + pU(w_0 - \alpha pqL - tqL, h_0 - (1 - r(\alpha + t))L)\}.$$

This extension can be seen as a more general version of the base model, which is nested for $r'' = 0$. As we can see in Proposition C.1 below, the underinsurance result of the base model holds in this extension, even though the condition on the marginal rate of substitution changes to accommodate the new health production function.

Proposition C.1 *In the extended model, if $qU_w(w_0, h_0) > r'(1)U_h(w_0, h_0)$ and insurance is actuarially fair, decision-makers do not purchase full insurance.*

Proof The proof follows the same steps as that of item (i) in Proposition 1. The set of possible choices is still a lattice. Further, $\partial^2 EU/\partial\alpha\partial{-t} = -pL[r'' U_h(y_{II}) + L(r')2U^{II}_{hh} + pLq(qU_{cc}(y_{II}) - r' U_{ch}(y_{II}))] > 0$, such that expected utility has increasing differences in $(\alpha, -t)$. Define $\hat{a}$ such that it fulfills $\partial EU \partial\alpha|_{t=0} = r'(\hat{a})U_h(w_0 - \hat{a}pqL, h_0 - (1 - r(\hat{a}))L) - q[(1 - p)U_c(w_0 - \hat{a}\text{pqL},\text{h}_0) + pU_c(w_0 - \hat{a}pqL, h_0 - (1 - r(\hat{a}))L)] = 0$. Using $U_{ccc} > 0$, we establish that $\partial^2 EU/\partial\alpha^2|_{t=0} < 0$ from $(1 - p)q^2U_{cc}(y_I) + pLq[qU_{cc}(y_{II}) - r' U_{ch}(y_{II})] + r'' U_h(y_{II}) + L(r')^2 U_{hh}(y_{II}) < 0$. Thus $\hat{a} < 1$ if $\partial EU/\partial\alpha|_{t=0,\alpha=1} < 0$. $\partial EU/\partial\alpha|_{t=0,\alpha=1} = pL[r'(1) U_h(w_0 - pqL, h_0) - qU_c(w_0 - pqL, h_0)]$. From $qU_c(w_0, h_0) > r'(1)U_h(w_0, h_0)$, we know that $r'(1)U_h(w_0, h_0) - qU_c(w_0, h_0) < 0$ and thus $\partial EU/\partial\alpha|_{t=0,\alpha=1} < 0$.

Further, from $r' U_{ch} > qU_{cc}$, $r'(1)U_h(w_0, h_0) - r'(1)U_h(w_0 - qpL, h_0) > qU_c(w_0, h_0) - qU_c(w_0 - pqL, h_0)$. Rearranging renders $\partial EU/\partial\alpha|_{\alpha=1} < r'(1)U_h(w_0, h_0) - qU_c(w_0, h_0) < 0$ and thus $\hat{a} < 1$. $\alpha^* < 1$ follows from the same steps as in Proposition 1.

Changing the health production function does not affect the result of Proposition 2, because the function is assumed to be independent of the decision-maker's wealth. I state this result as Proposition C.2 for completeness.

Proposition C.2 *Ceteris paribus, the expression $qU_c(w_0, h_0) - r(1)U_h(w_0, h_0)$ is decreasing in w_0.*