A theory of ex post moral hazard

Dario Gori* and Kourosh Khounsari†

September 12, 2026

Abstract

We study a monopoly health insurer when consumers privately know their illness risks and choose whether to obtain a preventive visit. The visit eliminates the health loss from illness without changing its probability, and the insurer cannot distinguish it from treatment after illness. Every optimal menu pools an interval of the highest-risk consumers induced to wait. The same contract can also be assigned to all higher-risk consumers, who instead visit preventively. Insurance reduces preventive use under symmetric information and never raises it above autarky when only the action is hidden. With both hidden type and hidden action, every nontrivial optimum raises preventive use and health expenditure under nondecreasing absolute risk aversion. The comparison can reverse under sufficiently strong decreasing absolute risk aversion and suitable type distributions. Under additional assumptions, separating tariff segments are concave below a threshold and convex above it; interior pools can generate upward kinks.

*dario.gori@tse-fr.eu

†seyed.khounsari@tse-fr.eu

The authors acknowledge funding from the French National Research Agency (ANR) under the Investments for the Future (Investissements d'Avenir) program, grant ANR-17-EURE-0010.

1 Introduction

Insurance affects both who bears the cost of medical care and how much care is used. The latter response is commonly referred to as ex post moral hazard (Pauly, 1968). Evidence from the RAND and Oregon experiments shows that insurance can increase medical use (Newhouse, 1993; Finkelstein et al., 2012). This evidence does not, however, imply that an optimal insurance contract must increase utilization relative to no insurance. In our model, a preventive visit is a form of self-insurance: it leaves the probability of illness unchanged, but eliminates the loss if illness occurs, so formal insurance can replace prevention.[1] With hidden action, full insurance violates the incentive to wait, and with hidden type, screening links the coverage offered to different risk types. The effect of insurance on preventive use can therefore reverse. This paper asks when insurance increases preventive care and health expenditure, and how the interaction between adverse selection and hidden action shapes the optimal menu of contracts.

We study a risk-averse consumer who privately observes his probability p of becoming ill before choosing a contract. The consumer can obtain a preventive visit at a cost d, which eliminates the health loss associated with illness, or wait until his health state is realized. If he waits and becomes ill, he must visit the doctor after the illness occurs. A contract consists of a premium t and an indemnity x paid whenever a visit takes place. The insurer observes the claim but cannot distinguish a preventive visit from a visit triggered by realized illness. Consequently, the insurer earns $t - px$ from a consumer who waits and $t - x$ from one who visits preventively. The type is private information, generating adverse selection, and the reason for the visit is unobservable, generating moral hazard. Higher-risk consumers both value insurance more and are more inclined to take the preventive action, so screening and action incentives cannot be analyzed separately.

[1] The medical-use margin in our model is a preventive visit chosen before illness is realized; treatment after illness is compulsory. Our use of ex post moral hazard refers to the potentially unnecessary utilization of preventive visits to insurance coverage.

In every information regime, behavior is characterized by an action cutoff: types below the cutoff wait, while types above it obtain the preventive visit. In autarky, this cutoff is p^A. Under symmetric information, the insurer provides full insurance to the types whom it induces to wait, and the cutoff satisfies $p^{SI} > p^A$. Insurance, therefore, replaces the preventive visit, which acts as self-insurance in our model, and lowers health expenditure. When the type is observable but the action is hidden, full insurance cannot induce waiting, because it makes the preventive visit strictly preferable. Nevertheless, the hidden-action cutoff satisfies $p^{HA} \in [p^A, p^{SI})$, so hidden action alone never raises preventive use above its autarky level. Once the type is also hidden, adverse selection can change this conclusion. Proposition 3 shows that, under nondecreasing absolute risk aversion, every nontrivial optimal menu satisfies $p^* < p^A$; in particular, under CARA or IARA, $p^* < p^A = p^{HA} < p^{SI}$. Since expenditure is decreasing in the action cutoff, these inequalities also rank health expenditure. Under sufficiently strong DARA and the distributional condition in Proposition 3, the comparison with autarky reverses and $p^A < p^* < p^{SI}$. Thus, hidden type and hidden action always generate more health expenditure than symmetric information, but whether they generate ex post moral hazard relative to autarky depends on wealth effects.

Our main result concerns the shape of the optimal menu. Theorem 2 shows that, at every optimum, the terminal net-coverage constraint binds and there exists $\bar{p} < p^*$ such that all types in $[\bar{p}, p^*]$ receive the same contract. This result requires neither a smooth density nor a shape restriction on the type distribution. Moreover, there is an outcome-equivalent representation in which the same contract is assigned to every type above $\bar{p}$: types in $[\bar{p}, p^*]$ wait, whereas types above p^* take the preventive visit. The action, therefore, changes at p^* even though neither the premium nor the indemnity changes. This contrasts with the standard monopoly-insurance problem with adverse selection, in which the highest type receives full insurance and pooling at the top is ruled out (Chade and Schlee, 2012).

Terminal bunching results from the interaction between the screening and action margins. As in a standard adverse-selection problem, the insurer would like to offer more coverage to higher-risk consumers. Incentive compatibility, however, transmits an increase in net coverage to all higher types. Once these types take the preventive visit, the insurer pays the indemnity with probability 1 rather than with probability p. Increasing coverage near the action cutoff, therefore, raises the cost of the positive mass of those who take the preventive visit and also changes which consumers choose the preventive action. At the optimum, the insurer stops differentiating contracts before reaching the cutoff.

We also characterize the tariff away from the terminal pool and show that hidden action changes familiar comparative statics. Under the regularity conditions in Proposition 5, a log-concave density and nonincreasing absolute risk aversion generate a global ordering of smooth curvature: every separating portion below a common threshold is concave, and every separating portion above it is convex. Interior pools bordered by separating regions produce upward kinks. This extends the backward-S characterization of Chade and Schlee (2012), which is derived under complete sorting, to menus that may contain pooling. Hidden action also overturns their value function comparative statics. Proposition 4 shows that under CARA, the insurer's profit converges to zero as either risk aversion or illness severity becomes arbitrarily large, while Example 1 shows that making a DARA consumer poorer can strictly reduce profit. These reversals arise because the changes that increase the demand for insurance can simultaneously make the preventive visit more attractive.

The paper contributes to the literature on insurance with endogenous medical use and private information. Zeckhauser (1970) studies the trade-off between risk sharing and incentives for medical use. Boone (2015) examines insurance design when ex post moral hazard is combined with adverse selection, using mean–variance preferences. We study the interaction between these two informational frictions under general risk-averse preferences, focusing on its implications for screening, preventive care, and

terminal bunching. Jullien et al. (1999) study how risk aversion affects self-insurance and self-protection, while Jullien et al. (2007) study screening when risk aversion is private information and effort is unobservable. Recent work studies how moral hazard and adverse selection can be disentangled and how multidimensional private information shapes health-insurance menus (Castro-Pires et al., 2024; Chade et al., 2022). Our closest benchmark is Chade and Schlee (2012), who characterize monopoly insurance under adverse selection alone. We add an unobservable preventive action, which rules out full insurance for consumers induced to wait and generates pooling below the action cutoff. Section 2 studies autarky, Section 3 solves the symmetric-information benchmark, Section 4 introduces hidden action with an observable type, and Section 5 analyzes the environment in which both the type and the action are private information.

2 Autarky

There exists a consumer whose health state is random: with probability $1-p$, the consumer is healthy and has wealth h; with probability p, he is ill and has wealth $i < h$. The probability p is distributed according to F, which has support $[0,1]$ and is continuous except for a possible atom at 1. Preferences are represented by a twice continuously differentiable Bernoulli utility u with $u' > 0$ and $u'' < 0$. The domain of u is either $\mathbb{R}$ or $(0,\infty)$; in the latter case, assume $i > d$. He privately observes his illness probability p before choosing an action $a \in \{v, w\}$. A preventive visit $(a = v)$ costs $d > 0$ and eliminates the health loss if illness occurs, yielding wealth $h - d$ in either state. Intuitively, the potential illness is still in an early stage. So the doctor can intervene successfully and ensure the agent's health does not worsen. Alternatively, the consumer waits $(a = w)$: he retains wealth h if healthy, but must obtain treatment after illness,[2] yielding wealth $i - d$. Thus the respective utilities are

[2]The underlying idea is that when the illness hits, the agent needs to see a doctor; otherwise, the utility loss becomes too extreme.

$u(h-d)$ and $(1-p)u(h)+pu(i-d)$. The preventive visit is complete self-insurance at cost d; it does not change the probability of illness.

The consumer takes the preventive visit if and only if $p > p^A$, where $p^A \in (0,1)$ is uniquely defined by

$$u(h-d) = p^A u(i-d) + (1-p^A)u(h). \tag{1}$$

We assume that the consumer waits when indifferent. Following this reasoning, we can define the expected health expenditure as $e^A = e(p^A)$, where

$$e(p) = \Big(1 - F(p) + \int_0^p \tilde{p}\,\mathrm{d}F(\tilde{p})\Big)d.$$

Note that $e(p)$ is strictly decreasing in p, so higher cutoffs mean lower expenditure. Before moving to the next section, let us also define the agent's indirect utility in autarky as

$$U_0(p) = \max\{u(h-d), pu(i-d)+(1-p)u(h)\}.$$

3 Symmetric Information

Let us now study the problem of a monopolist selling insurance to the consumer. For this and the following sections, the timing is the following:

1. The principal offers a menu of contracts.
2. p is realized, and the agent picks at most one contract from the menu.
3. The agent chooses $a \in \{v, w\}$.
4. Payoffs are realized.

Let us first focus on the case of symmetric information, where the monopolist observes both the action taken by the agent, a, and the type of the agent, p. This has two consequences. First, each contract can be contingent on p. Moreover, the contract

determines the agent's actions. So, the menu can be represented by the following set $\mathcal{M}^{SI} = \{(t(p), x(p), a(p)) \mid p \in [0,1]\}$. Let us define the utility that a type p gets from accepting a contract as

$$U^{SI}(p) = \begin{cases} u(h-d-t(p)+x(p)) & \text{if } a(p) = v, \\ pu(i-d-t(p)+x(p)) + (1-p)u(h-t(p)) & \text{if } a(p) = w. \end{cases}$$

The monopolist's objective is to maximize its expected profit. Therefore, we can write the principal's problem as follows

$$\begin{aligned} \max_{\mathcal{M}^{SI}} \ & \mathbb{E}[t(p) - px(p)\mathbb{1}_{\{a(p)=w\}} - x(p)\mathbb{1}_{\{a(p)=v\}}] \\ \text{s.t. } & U^{SI}(p) \geq U_0(p) \qquad \forall p \in [0,1]. \end{aligned}$$

Notice that the principal's problem is equivalent to a continuum of pointwise maximization problems since the contracts are conditioned on p. Inducing a preventive visit yields no positive profit.[3] Inducing waiting is cheapest under full insurance, with

$$x(p) = h - i + d, \qquad t(p) = h - u^{-1}(U_0(p)).$$

This gives a nonnegative profit to the monopolist if $t(p) \geq p\,(h-i+d)$, meaning the premium is at least as high as the fair premium. Equivalently:

$$U_0(p) \leq u\big((1-p)h + p(i-d)\big). \tag{2}$$

Strict concavity makes this inequality strict for $p \in (0, p^A]$, because

$$U_0(p) = pu(i-d) + (1-p)u(h) < u(p(i-d) + (1-p)h).$$

For $p \geq p^A$, eq. (2) is equivalent to $p \leq p^{SI} \equiv d/(h-i+d)$. Hence, the insurer assigns

[3] For any p, $U^{SI}(p) = U_0(p)$; otherwise, the principal could increase the premium. Therefore, if $a(p) = v$, then

$$u(h-d-t(p)+x(p)) = U^{SI}(p) = U_0(p) \geq u(h-d),$$

meaning $x(p) \geq t(p)$. So, the profit the monopolist can achieve if $a(p) = v$ is bounded above by 0.

full insurance and waiting below p^{SI}, and the zero contract and a preventive visit above it. In particular, $p^A < p^{SI}$ and $e^{SI} = e(p^{SI}) < e^A$: formal insurance displaces preventive self-insurance. Figure 1 illustrates the allocation.

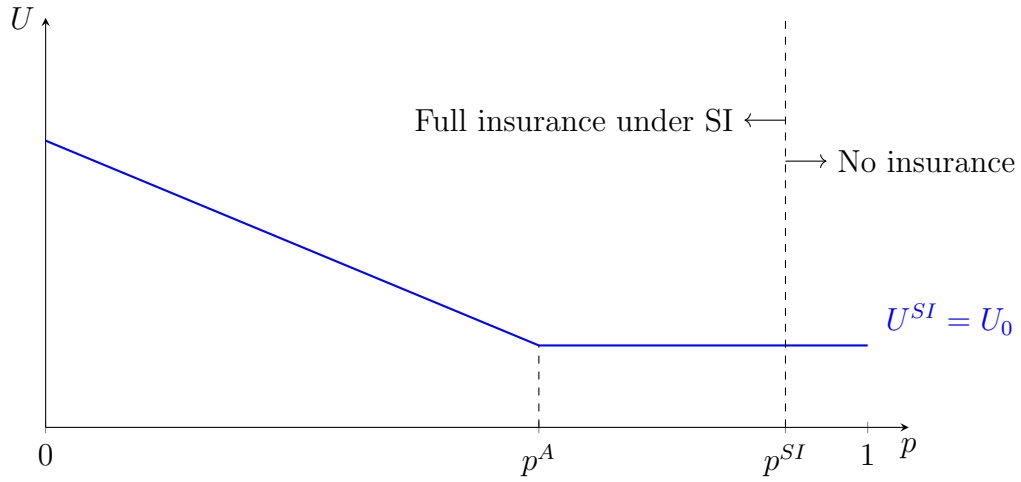

Figure 1: Symmetric Information

4 Hidden Action

In this benchmark, the principal does not observe the action $a \in \{v, w\}$ taken by the agent. So, the menu of contracts offered by the principal is $\mathcal{M}_{HA} = \{(t(p), x(p)) \mid p \in [0, 1]\}$. The problem of the principal can be written as:

$$\max_{\mathcal{M}^{HA}} \mathbb{E}[t(p) - px(p)\mathbb{1}_{\{a^*(p)=w\}} - x(p)\mathbb{1}_{\{a^*(p)=v\}}]$$
$$\text{s.t. } U^{HA}(p) \geq U_0(p),$$

where $a^*(p)$ is the best response of the agent and

$$U^{HA}(p) = \max\{u(h - d - t(p) + x(p)),\ pu(i - d - t(p) + x(p)) + (1 - p)u(h - t(p))\}.$$

Once again, the principal can maximize profit from each type separately since p is observable, and inducing a preventive visit yields no positive profit. We show that

there exists a threshold $p^{HA} \in [p^A, p^{SI})$ such that the profit of the monopolist from inducing $a^*(p) = w$ is strictly positive for $p \in (0, p^{HA})$ and is negative above this threshold. This means that with the optimal contract:

$$a^*(p) = \begin{cases} w & p \in (0, p^{HA}), \\ v & p \in (p^{HA}, 1]. \end{cases} \tag{3}$$

Let us focus on the profit when inducing $a^*(p) = w$. Fixing a p, the problem of the principal is the following:

$$\begin{aligned} \Pi^w(p) \equiv \sup_{t,x} \quad & t - px \\ \text{s.t.} \quad (\text{PC}) \quad & pu(i - d - t + x) + (1 - p)u(h - t) \geq U_0(p) \\ (\text{IC}) \quad & pu(i - d - t + x) + (1 - p)u(h - t) \geq u(h - d - t + x). \end{aligned}$$

We use the convention $\sup \varnothing = -\infty$, since for some types no contract can induce waiting.

The following proposition establishes the threshold structure of the wait-inducing profit. All missing proofs are in the appendix.

Proposition 1. *There exists a cutoff*

$$p^{HA} \in [p^A, p^{SI})$$

such that

$$\Pi^w(p) > 0 \qquad \text{for } p \in (0, p^{HA}),$$

and

$$\Pi^w(p) < 0 \qquad \text{for } p > p^{HA}.$$

At the boundary points $p = 0$ and at $p = p^{HA}$, the profit is zero.

For $p \in (0, p^A)$, a small amount of actuarially fair insurance makes participation strictly attractive while preserving the incentive to wait, so the insurer can charge a

positive loading. For $p \geq p^A$, the outside option is constant. Any nonnegative-profit contract feasible for a higher type is feasible and strictly more profitable for a lower type in this range. Finally, strict concavity rules out nonnegative wait-inducing profit at p^{SI}. These observations yield the cutoff.

Since any contract inducing v gives the principal weakly negative profit, Proposition 1 implies that, away from the boundary cases, the optimal action has the threshold form described in eq. (3). The boundary types have probability zero under F, so their assigned actions do not affect the expected profit and are left unspecified. The next theorem characterizes the optimal insurance plan for $p \in (0, p^{HA})$.

Theorem 1. *For $p \in (0, p^{HA})$, IC binds at the optimum. In addition, either PC binds or*

$$\frac{u'(h-d-t+x)}{u'(i-d-t+x)-u'(h-t)} = p. \tag{4}$$

Suppose absolute risk aversion is monotone and define

$$\Gamma \equiv \frac{u'(h-d)}{u'(i-d)-u'(h)}.$$

If $p^A \leq \Gamma$ (which is true for all IARA, CARA, and some DARA utility functions), the PC is always binding and $p^{HA} = p^A$. Conversely, if $p^A > \Gamma$, there exists a threshold $\tilde{p} \in (0, p^A)$ such that PC binds for $p \in (0, \tilde{p}]$ and is slack for $p \in (\tilde{p}, p^{HA})$. In this PC-slack region, the agent's utility, $U^{HA}(p)$, is strictly increasing in p.

Since $p^{HA} \geq p^A$, hidden action alone weakly reduces expenditure relative to autarky: $e^{HA} = e(p^{HA}) \leq e^A$. Binding IC requires $x < d$, so the insurer provides partial insurance. When PC also binds, the consumer receives exactly his outside option. Under CARA or IARA, this occurs for every served type and $p^{HA} = p^A$.

With sufficiently strong DARA, leaving rent can instead reduce the cost of inducing waiting. Raising the premium makes the consumer poorer and more risk-averse, requiring additional coverage to preserve IC. Equation (4) equates the resulting marginal expected cost to the marginal premium revenue. The condition $p^A > \Gamma$

identifies when this effect is strong enough to make rent profitable at the autarky cutoff. PC then becomes slack above $\tilde{p}$, the waiting cutoff increases to $p^{HA} > p^A$, and the utility promised to served types increases with p throughout the PC-slack region. Figures 2 and 3 summarize these results.

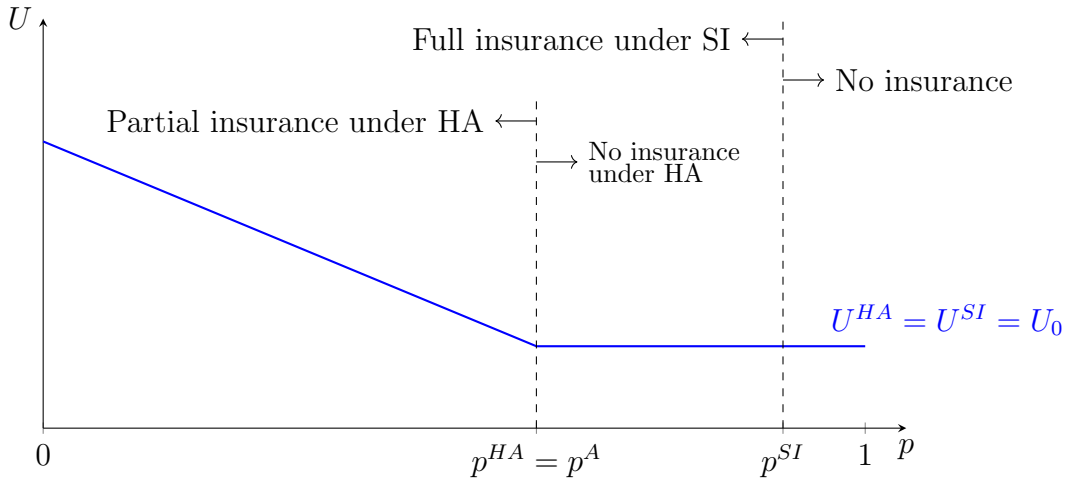

Figure 2: Hidden Action for $p^A \leq \Gamma$ (including CARA or IARA utility functions)

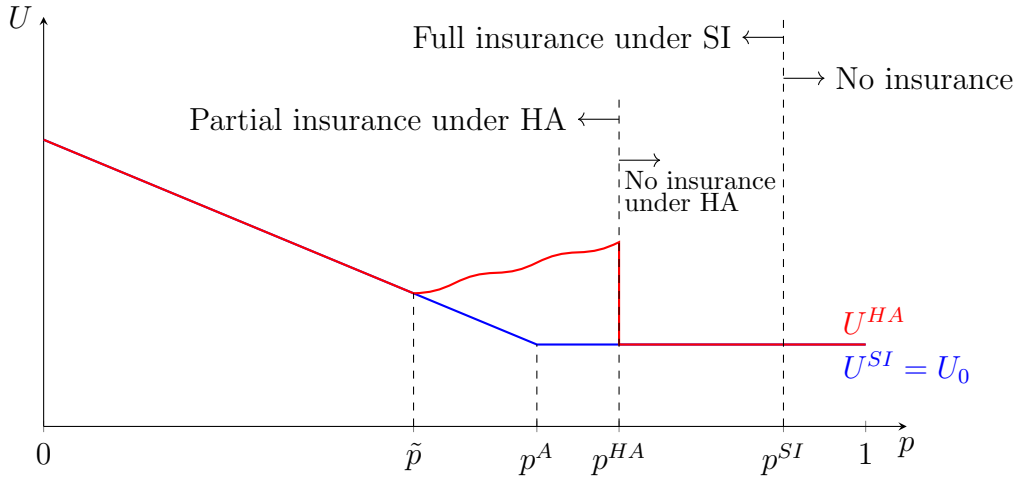

Figure 3: Hidden Action for $p^A > \Gamma$ (some DARA utility functions)

With private types, truthfulness requires the utility of wait types to decrease with risk. The rent-leaving allocation under strong DARA, therefore, cannot be

implemented unchanged. More generally, screening links the contracts of different types: higher net coverage for wait types also increases the insurer's losses on visit types. This link drives terminal pooling in the next section.

5 Hidden Type and Action

We now analyze the problem of a monopolist who observes neither the agent's risk type nor his action. We can directly apply the revelation principle and focus on the same class of menus of contracts as in the previous section: $\mathcal{M} = \{(t(p), x(p))\}$. The difference from the previous problem, of course, is that the agent can pick the contract he prefers from the menu. So, let us define the indirect utility that type p derives from choosing the contract designed for type $\hat{p}$ as

$$U(\hat{p}, p) = \max\{u(h - d - t(\hat{p}) + x(\hat{p})),\ pu(i - d - t(\hat{p}) + x(\hat{p})) + (1 - p)u(h - t(\hat{p}))\},$$

and the indirect utility associated with his contract as $U(p) = U(p, p)$.
We say that a menu $\mathcal{M}$ is **truthful** if

$$U(p) = \max_{\hat{p}} U(\hat{p}, p).$$

For any incentive-compatible menu $\mathcal{M}$, we can identify two subsets of types: the types that are induced to avoid a preventive visit and wait,

$$\mathcal{W} = \{p \in [0, 1] :\ U(p) = pu(i - d - t(p) + x(p)) + (1 - p)u(h - t(p))\},$$

and the ones that decide to make a preventive visit, $\mathcal{V} = [0, 1] \setminus \mathcal{W}$. As in Chade and Schlee (2012), we reformulate the problem as one in which the principal chooses a menu of state-contingent utilities rather than a menu of premium–indemnity pairs. Given a menu $\mathcal{M}$, we define $u_h(p) \equiv u(h - t(p))$, $u_i(p) \equiv u(i - d - t(p) + x(p))$ and $\Delta(p) \equiv u_h(p) - u_i(p)$. So, we can rewrite $U(p) = u_h(p) - p\Delta(p)$ for all $p \in \mathcal{W}$. Notice that there is a one-to-one mapping between the part of the menu designed for $\mathcal{W}$ and

the set $\{(u_h(p), \Delta(p))\}$. Indeed, denoting $g(\cdot) = u^{-1}(\cdot)$, we obtain

$$t(p) = h - g(u_h(p)),\ x(p) = g(u_h(p) - \Delta(p)) - g(u_h(p)) + h + d - i. \tag{5}$$

Before turning to the monopolist's problem, we characterize truthful menus in the following Lemma. For a type who visits, only net coverage matters: the agent obtains $u(h - d + x - t)$, and the principal earns $t - x$. We henceforth use *canonical* menus: the wait set is nonempty and closed, and all visit types pay a common premium. Every truthful menu has a canonical representative with the same indirect utilities and expected profit; the construction is given in the proof of Lemma 1.

Lemma 1. *A canonical menu $\mathcal{M}$ is truthful if and only if*

1. *there exists $p^* \in [0, 1)$ such that $\mathcal{W} = [0, p^*]$, $\mathcal{V} = (p^*, 1]$;*
2. *$\Delta(\cdot)$ is nonincreasing on $\mathcal{W}$, and $U(\cdot)$ is Lipschitz continuous, nonincreasing, and convex with*
   - *$\dot{U}(p) = -\Delta(p)$ for a.e. $p \in \mathcal{W}$;*
   - *$U(p) = u(h - d + k)$ for all $p \in \mathcal{V}$, where $k = \sup_{p \in [0,1]} \{x(p) - t(p)\}$.*

Truthfulness also implies that the premium and net coverage are nondecreasing in risk.

Lemma 2. *If $\mathcal{M}$ is a truthful canonical menu, then $u_h(\cdot)$ is nonincreasing and $u_i(\cdot)$ is nondecreasing; equivalently, the premium $t(\cdot)$ and net coverage $x(\cdot) - t(\cdot)$ are nondecreasing. Consequently, $x(\cdot)$ is also nondecreasing.*

Let $E_0^{p^*} = \int_0^{p^*} p \, \mathrm{d}F(p)$. For $p \leq p^*$, the envelope condition implies

$$U(p) = U(0) - \int_0^p \Delta(s) \, \mathrm{d}s,$$

whereas $U(p) = U(p^*)$ for $p > p^*$. The principal's problem can therefore be written

as

$$\max_{p^*, U(0), \Delta(\cdot)} \quad (1 - E_0^{p^*})h + E_0^{p^*} i - (1 - F(p^*) + E_0^{p^*})d - (1 - F(p^*))g(U(p^*))$$
$$- \int_0^{p^*} \Big[p\ g(U(p) - (1-p)\Delta(p)) + (1-p)g(U(p) + p\Delta(p)) \Big] \, \mathrm{d}F(p)$$

s.t.

$$\Delta(\cdot) \geq 0 \text{ nonincreasing},$$
$$\dot{U}(p) = -\Delta(p) \qquad \text{for a.e. } p \in [0, p^*],$$
$$U(p) \geq U_0(p) \qquad \text{for every } p \in [0, 1],$$
$$u(g(U(p^*)) - h + i) - U(p^*) + (1 - p^*)\Delta(p^*) \geq 0.$$

The objective follows by substituting $u_h = U + p\Delta$ and $u_i = U - (1-p)\Delta$ into eq. (5) for wait types. For visit types, continuity gives $U(p^*) = u(h - d + k)$, so profit is $h - d - g(U(p^*))$.

The first two constraints are the local incentive-compatibility restrictions from Lemma 1. The third constraint is the participation constraint. The final constraint ensures that the net coverage assigned to visit types is at least as large as the net coverage assigned to the highest wait type. Indeed, since $g(U(p^*)) = h - d + k$, the final constraint is equivalent to

$$u(i - d + k) \geq U(p^*) - (1 - p^*)\Delta(p^*) = u_i(p^*).$$

Since u is increasing, this is equivalent to $k \geq x(p^*) - t(p^*)$. By Lemma 2, net coverage $x(p) - t(p)$ is nondecreasing on $\mathcal{W}$, so this single terminal constraint implies

$$k \geq x(p) - t(p) \qquad \text{for every } p \in \mathcal{W}.$$

Together with $x(p) - t(p) = k$ on $\mathcal{V}$, it ensures that k is the maximum net coverage in the menu.

Conversely, every feasible choice in this program is implementable. Write $U^* =$

$U(p^*)$ and $k = g(U^*) - h + d$. Reconstruct the wait contracts using Equation (5), and assign $(d-k, d)$ to types above p^*. The terminal constraint implies

$$(1-p^*)\Delta(p^*) \geq U^* - u\big(g(U^*) - h + i\big) > 0.$$

Thus $\Delta > 0$ throughout the wait region. Monotonicity and the envelope formula imply that each assigned wait payoff supports U from below and that net coverage is nondecreasing there. Its terminal value is at most k, so no contract offers a visit payoff above U^*. The common visit contract offers utility U^*, while every wait payoff is strictly below U^* above p^*. Hence the assigned actions are optimal and the resulting menu is truthful by lemma 1. The following lemma shows that the participation constraints can be reduced once the utility spread is bounded.

Lemma 3. *Let M be a truthful canonical menu. If*

$$U(0) \geq u(h) \qquad \text{and} \qquad 0 \leq \Delta(p) \leq u(h) - u(i-d) \quad \text{for every } p \in \mathcal{W},$$

then

$$U(p) \geq U_0(p) \qquad \text{for every } p \in [0,1].$$

We next establish that these restrictions can be imposed without loss and that the principal's problem has a solution.

Proposition 2. *The principal's problem admits an optimal truthful canonical menu. Every optimal truthful canonical menu satisfies*

$$U(0) = u(h) \qquad \text{and} \qquad 0 \leq \Delta(p) \leq u(h) - u(i-d) \quad \text{for every } p \in \mathcal{W} \setminus \{0\}.$$

Moreover, its action cutoff is interior:

$$p^* \in (0,1).$$

Normalize type 0's contract to $(0,0)$, which preserves utilities and expected profit. By Lemma 3 and Proposition 2, we may then replace the participation constraints

with

$$U(0) = u(h), \qquad 0 \leq \Delta(p) \leq u(h) - u(i-d) \quad (p \in \mathcal{W}).$$

We now study the upper end of the menu.

In the standard monopoly-insurance problem with adverse selection, the highest-risk type receives full insurance, and there is no pooling at the top of the menu (Chade and Schlee, 2012). Hidden action changes this conclusion. Full insurance cannot be assigned to a type who is induced to wait, because it would make the preventive visit strictly preferable. Instead, the moral-hazard constraint generates a pooling interval immediately below the action cutoff.

Theorem 2. *Consider an optimal truthful canonical menu. Then the terminal net-coverage constraint binds:*

$$k = x(p^*) - t(p^*).$$

Moreover, there exists $\bar{p} \in [0, p^*)$ *such that*

$$\Delta(p) = \Delta(p^*) \qquad \textit{for every } p \in [\bar{p}, p^*].$$

Consequently,

$$t(p) = t(p^*) \qquad \textit{and} \qquad x(p) = x(p^*)$$

for every $p \in [\bar{p}, p^*]$*: the highest wait types are pooled at the same contract.*

The proof compares the optimal menu with one that retains the contracts below a wait type q and assigns q's contract to all higher types. This truncation preserves truthfulness and participation. If the terminal constraint were slack, taking $q = p^*$ would immediately reduce the losses on visit types without affecting lower types. If contracts continued to differ arbitrarily close to p^*, taking q just below the cutoff would reduce net coverage for every original visit type, while any loss on wait types would be confined to a shrinking interval. The gains from the visit types dominate the costs imposed on the affected wait types, so the highest wait types must be pooled.

The pooling result has a distinctive interpretation. Since the terminal constraint

binds, the visit utility generated by the terminal wait contract is

$$u\big(h - d + x(p^*) - t(p^*)\big) = u(h - d + k) = U(p^*).$$

Thus, type p^* is indifferent between waiting and taking the preventive visit under this contract. Types just below p^* put less weight on the sick state and choose to wait, while types above p^* put more weight on the sick state and choose the preventive visit. Therefore, the action changes at p^* even though the contract need not change. In this representation, the expected expenditure per consumer jumps from p^*d to d at the action cutoff, even though the premium and indemnity remain unchanged.

The terminal-bunching result also allows us to compare the action cutoff under hidden types with the autarky cutoff p^A. The comparison depends on the agent's wealth effects. Under CARA or IARA, the terminal insurance contract makes waiting less attractive than in autarky. The terminal contract weakly raises visit and sick-state wealth and reduces the healthy-state advantage of waiting. Under CARA or IARA, both effects make the preventive visit attractive at a lower risk than in autarky. Hence, the switch to the preventive visit occurs strictly before p^A whenever the optimal menu provides nonzero insurance. Under DARA, the wealth effect works in the opposite direction. In that case, insurance may make the agent sufficiently less risk-averse that waiting remains attractive beyond p^A. Whether this occurs also depends on the distribution of types. The next proposition gives a primitive sufficient condition.

Proposition 3. *The optimal action cutoff satisfies $p^* < p^{SI}$. Moreover:*

1. *If absolute risk aversion is nondecreasing, then*

$$p^* < p^A,$$

whenever the allocation is not outcome-equivalent to no insurance. Equality $p^* = p^A$ occurs if the optimal allocation is outcome-equivalent to no insurance.[4]

[4] Under CARA utility $u(c) = -e^{-\alpha c}$, the zero contract is optimal if $F(p^A) \leq e^{-\alpha d}$; see the proof

2. Suppose that absolute risk aversion is nonincreasing and

$$p^A > \Gamma \equiv \frac{u'(h-d)}{u'(i-d)-u'(h)}.$$

Then $p^{HA} > p^A$. Fix any $p_0 \in (p^A, p^{HA})$, and let $\pi_0 \equiv \Pi^w(p_0) > 0$ denote the maximal pointwise profit from inducing type p_0 to wait. If there exists $\varepsilon \in (0, \frac{\pi_0}{2(\pi_0+d)})$ such that $F(p^A) \leq \varepsilon$ and $1 - F(p_0) \leq \varepsilon$ then

$$p^* > p^A.$$

The inequality $p^* < p^{SI}$ implies that expenditure under hidden type and action, $e^* \equiv e(p^*)$, is strictly greater than under symmetric information. Part 1 shows that, under CARA or IARA, private information reinforces the use of the preventive visit. The action cutoff lies below its autarky level: insurance is supplied to some low-risk types, but the need to screen types makes higher-risk agents switch to the safe preventive visit earlier than they would in autarky. The inequality is strict whenever the insurer offers a nontrivial menu. Consequently, under CARA or IARA, $e^* \geq e^A$, with strict inequality whenever the optimal allocation is not outcome-equivalent to no insurance.

Part 2 gives sufficient conditions for the opposite ordering under DARA. The condition $p^A > \Gamma$ makes it profitable to induce some types above p^A to wait. If enough probability mass lies in $(p^A, p_0]$ and little lies in either tail, these profits outweigh the losses on visit types, and the optimal wait region extends beyond p^A. Insurance then reduces expenditure relative to autarky, so $e^{SI} < e^* < e^A$.

5.1 Wealth, risk aversion, and illness severity

Chade and Schlee (2012) show that, in the standard monopoly-insurance problem

of proposition 3. Intuitively, every nontrivial menu must provide positive net coverage for the types who take the preventive visit and therefore incur a loss on each of them. If the mass of types below p^A is sufficiently small, the profits obtained from the types who wait cannot compensate for these losses. In this case, the optimal allocation is outcome-equivalent to no insurance, and hence $p^* = p^A$.

with adverse selection, the principal's maximum profit increases with the size of the loss and with the agent's risk aversion, and decreases with the agent's wealth under DARA. These conclusions do not generally extend to our environment because the same changes also affect the agent's decision to wait or take the preventive visit. To make the comparison transparent, write $h = w$ and $i = w - D$, where w is initial wealth and $D = h - i > 0$ is the health loss avoided by the preventive visit. Holding d and F fixed, let $P(u, w, D)$ denote the principal's maximum expected profit.

Proposition 4. *Suppose $u_\alpha(c) = -e^{-\alpha c}$ for $c \in \mathbb{R}$, with $\alpha > 0$. Then*

$$0 \leq P(u_\alpha, w, D) \leq d\, F(p^A).$$

Consequently,

$$\lim_{\alpha \to \infty} P(u_\alpha, w, D) = 0 \qquad \text{and} \qquad \lim_{D \to \infty} P(u_\alpha, w, D) = 0.$$

Proposition 4 does not say that profit is globally decreasing in risk aversion or illness severity. It shows instead that the monotonicity results of Chade and Schlee (2012) cannot hold globally: whenever profit is positive at some finite value of α or D, a sufficiently large increase in that primitive eventually lowers profit. The mechanism is the endogenous action choice. A larger loss or greater risk aversion raises the demand for insurance, but it also makes the safe preventive visit more attractive. If a type switches from waiting to visiting under a contract (t, x), the principal's profit falls from $t - px$ to $t - x$, a reduction of $(1 - p)x$. Under CARA, the set of types who can profitably be induced to wait shrinks toward zero as either α or D becomes large.

The wealth comparative-static results under DARA can also reverse.

Example 1. *Let $u(c) = \log c$ and $D = d = 1$. There exists a smooth density f, strictly positive on $[0, 1]$, such that*

$$P(u, 4, D) > P(u, 3, D).$$

Thus, making the DARA agent poorer can strictly lower the principal's maximum

profit.

The proof concentrates a smooth, full-support density near $p_b = 2/5$. This type can be profitably induced to wait when $w = 4$, but not when $w = 3$, because $p^A(3) < p_b < p^A(4)$ and $p^{HA}(w) = p^A(w)$ at both wealth levels. Together with proposition 4, this shows that all three value-function comparative statics can reverse when the insured can take a preventive action.

5.2 Curvature of the premium schedule

We now study the premium as a function of coverage. By Lemma 2, both $t(\cdot)$ and $x(\cdot)$ are nondecreasing. Moreover, if two types receive the same coverage, monotonicity of the premium and net coverage implies that they also pay the same premium. The menu therefore induces a nondecreasing premium schedule T on its coverage range, with $t(p) = T(x(p))$.

Assume throughout this subsection that u is three times continuously differentiable, that F admits a strictly positive, continuously differentiable, log-concave density f on $(0, p^*)$, and that absolute risk aversion $A(c) \equiv -u''(c)/u'(c)$ is nonincreasing. We also assume that there exists a finite partition of $[0, p^*]$ such that Δ is continuously differentiable on the interior of each element of the partition. A differentiability point is separating if $\Delta'(p) < 0$, and an interval is pooling if Δ is constant on it. Let $I(p) \equiv i - d - t(p) + x(p)$ and $H(p) \equiv h - t(p)$.

Proposition 5. *The optimal contract schedule $p \mapsto (t(p), x(p))$ is continuous on $(0, p^*)$, and its induced premium schedule T is continuous on $x((0, p^*))$. At every separating type,*

$$T'(x(p)) = \frac{p\,u'(I(p))}{(1-p)u'(H(p)) + p\,u'(I(p))}, \qquad p < T'(x(p)) < 1.$$

Moreover, there exists $\widehat{p} \in [0, p^*]$ such that, at every separating type,

$$T''(x(p)) < 0 \quad \text{if } p < \widehat{p}, \qquad T''(x(p)) > 0 \quad \text{if } p > \widehat{p}.$$

At $p = \widehat{p}$, curvature may be zero or the threshold may lie inside a pooling interval. Hence at most one separating component contains a change from concavity to convexity.

If $[p_1, p_2] \subset (0, p^*)$ is a pooling interval with separating regions on both sides and common contract (t_0, x_0), then

$$T'_-(x_0) < T'_+(x_0).$$

Thus, every interior pool generates an upward kink in the premium schedule.

The proposition gives a global ordering of smooth curvature even when the menu is not completely separating. All separating portions below $\widehat{p}$ are concave, and all separating portions above it are convex. Pooling intervals generate upward kinks and may occur on either side of $\widehat{p}$. Thus, the whole tariff need not be globally concave below $\widehat{p}$ in the generalized sense. What the proposition rules out is a concave separating portion after smooth curvature has become convex. If there is no interior pooling below the terminal pool, smooth curvature can change only from concavity to convexity before the schedule reaches the pooled contract in theorem 2; either curvature region may be empty. By contrast, Chade and Schlee (2012) obtains a globally backward-S-shaped tariff after imposing complete sorting.

6 Conclusion

We study a monopoly health insurer facing both private information about the consumer's probability of illness and an unobservable decision to obtain a preventive visit. The insurer observes a medical claim but cannot distinguish a preventive visit from treatment after illness. The benchmark analysis shows that insurance need not increase medical use when preventive care provides self-insurance.

Under symmetric information, full insurance crowds out the preventive visit, which acts as self-insurance, and $p^{SI} > p^A$. With hidden action and an observable type,

$p^{HA} \in [p^A, p^{SI})$, so preventive use remains weakly below its autarky level. Once both the type and the action are hidden, the cutoff remains below p^{SI}, but its comparison with autarky depends on wealth effects. Under CARA or IARA, every nontrivial optimal menu satisfies $p^* < p^A = p^{HA}$, so adverse selection increases preventive visits and health expenditure relative to autarky. Under sufficiently strong DARA and the distributional condition in Proposition 3, the ordering can reverse and $p^* > p^A$. Thus, the interaction between adverse selection and hidden action always raises expenditure relative to symmetric information, but need not raise it relative to no insurance.

Our main result is terminal bunching. Whenever the optimal action cutoff is interior, the terminal net-coverage constraint binds and an interval of the highest types who wait receives a common contract. In an outcome-equivalent representation, this same contract is assigned to all higher types: those below p^* wait, whereas those above p^* take the preventive visit. Behavior and health expenditure therefore change discontinuously at the cutoff even though the premium and indemnity do not. The insurer's expected indemnity cost rises by $(1 - p^*)x(p^*)$, which is strictly positive whenever the terminal contract provides insurance. This pooling does not rely on the shape of the type distribution. It is generated by the interaction between screening and hidden action: differentiating contracts near the cutoff raises the net coverage that must be made available to higher types, including those for whom the indemnity is paid with probability one. This contrasts with the adverse-selection-only benchmark, in which the highest type receives full insurance and there is no pooling at the top (Chade and Schlee, 2012).

The same interaction changes the tariff and the insurer's comparative statics. Under the conditions of Proposition 5, all separating portions below a common threshold are concave, and all separating portions above it are convex, while the interior pools bordered by separating regions generate upward kinks. Moreover, changes that increase insurance demand in the standard model can make the preventive visit sufficiently attractive to reduce the insurer's profit: under CARA, profit converges to zero

as either risk aversion or illness severity becomes arbitrarily large, and under DARA, a poorer consumer can be less profitable.

Appendix

Proof of Proposition 1. Write $\pi = t - px$, $I = i - d - t + x$, $H = h - t$, and $V = h - d - t + x$. For $p \in (0, p^A)$, a sufficiently small actuarially fair indemnity strictly improves wait utility and preserves strict IC. A small positive loading therefore gives $\Pi^w(p) > 0$.

We record bounds that also establish the needed compactness. For $p \in (0, 1)$, let

$$L_p = pu'(i-d) + (1-p)u'(h), \qquad D_p = p(1-p)\big[u'(i-d) - u'(h)\big] > 0.$$

The supporting-line inequalities for u, together with PC, give

$$U_0(p) \le pu(I) + (1-p)u(H) \le pu(i-d) + (1-p)u(h) - L_p\pi + D_p x,$$

and hence $D_p x \ge L_p \pi$. Also, IC implies $I < V < H$, so $x < d$. Consequently, a nonzero feasible contract with $\pi \ge 0$ has $x > 0$ and $t > 0$. It must also have $t < x$: otherwise both wait-state wealth levels are weakly below their autarky levels, and one is strictly below, violating PC. Thus every nonnegative-profit contract belongs to the compact triangle

$$T = \{(t, x) : 0 \le t \le x \le d\}.$$

The zero contract is the only exception to the strict inequalities $0 < t < x < d$.

Let N be the set of types admitting a feasible nonnegative-profit contract. The preceding bound, and the direct observation at $p = 0$, allow N to be written as the projection of a closed subset of $[0, 1] \times T$; hence N is compact. It contains $[0, p^A]$. For $p \ge p^{SI}$, a feasible contract with $\pi \ge 0$ would satisfy

$$pu(I) + (1-p)u(H) < u\big(h - p(h - i + d) - \pi\big) \le u(h-d) = U_0(p),$$

contradicting PC. Here the strict inequality follows from $I < H$; at $p = 1$, IC is infeasible directly. Therefore

$$p^{HA} := \max N \in [p^A, p^{SI}).$$

If $p_2 > p_1 \geq p^A$ and $p_2 \in N$, a nonnegative-profit contract for p_2 is nonzero and has $x > 0$ and $H > I$. The same contract satisfies PC and IC at p_1 and raises profit by $(p_2 - p_1)x > 0$. This proves strict positivity below p^{HA}. Profit at p^{HA} is zero: if it were positive, reducing both t and x by a sufficiently small common amount would keep profit positive and make both constraints strict, since I and V remain unchanged while H rises. By continuity, the modified contract would remain feasible and profitable for a slightly higher type, contradicting the definition of p^{HA}.

Finally, for a fixed $p \in (0, 1)$, a feasible sequence with profits converging to zero satisfies $x < d$ and $x \geq L_p\pi/D_p$. It therefore has a subsequence converging to a contract in T with zero profit, and that limit is feasible. Thus outside N the supremum cannot equal zero, proving $\Pi^w(p) < 0$ for $p > p^{HA}$. The nonnegative optimum is attained by compactness whenever it exists. At $p = 0$, PC requires $t \leq 0$, so $\Pi^w(0) = 0$. □

Proof of Theorem 1. Fix $p \in (0, p^{HA})$. Proposition 1 and compactness ensure an optimal contract with positive profit. Write $\ell = h - i$, $I = V - \ell$, and $H = h - t$, where $V = h - d - t + x$. IC implies $I < V < H$. If IC were slack, increasing V by ε and reducing H by $p\varepsilon/(1-p)$ would preserve profit and strictly increase wait utility. A sufficiently small common reduction of both wealth levels would then increase profit while preserving PC and IC, a contradiction. Thus IC binds.

Let $g = u^{-1}$ and write $z = u(V)$. Binding IC determines

$$b(z) = u(g(z) - \ell), \qquad q(z, p) = \frac{z - pb(z)}{1 - p} = u(H).$$

The insurer therefore minimizes

$$K(z,p) = pg(z) + (1-p)g(q(z,p)) \quad \text{subject to} \quad z \geq U_0(p),$$

over admissible wealth levels, since profit is $h - pd - K(z,p)$. At a slack-PC optimum,

$$0 = K_z = pg'(z) + g'(q)\big(1 - pb'(z)\big).$$

Substituting $b'(z) = u'(I)/u'(V)$ gives $u'(V) = p[u'(I) - u'(H)]$, which is eq. (4).

First suppose $A = -u''/u'$ is nonincreasing. Then

$$b'(z) = \frac{u'(I)}{u'(V)} > 1, \qquad b''(z) = \frac{u'(I)}{[u'(V)]^2}\big[A(V) - A(I)\big] \leq 0.$$

Thus $q(\cdot,p)$ is convex and g is increasing and strictly convex, so $K(\cdot,p)$ is strictly convex. Its admissible domain is an interval: the requirement that $g(z) - \ell$ belong to the wealth domain defines an interval, while the remaining restriction is the sublevel condition $q(z,p) < \sup u$, since $q(z,p) > z$. More explicitly,

$$K_{zz} = pg''(z) + \frac{g''(q)(1-pb'(z))^2}{1-p} - pg'(q)b''(z) > 0.$$

Moreover,

$$q_p = \frac{z - b(z)}{(1-p)^2} > 0, \qquad K_{zp} = g'(z) - b'(z)g'(q) + (1 - pb'(z))g''(q)q_p.$$

At any zero of K_z, the last coefficient satisfies $1 - pb'(z) = -pg'(z)/g'(q) < 0$. Since $q > z$ and $b' > 1$, it follows that $K_{zp} < 0$ at every such zero.

For $p \leq p^A$, the PC boundary is admissible: setting $z = U_0(p)$ gives $g(z) \geq h - d$ and $q(z,p) \leq u(h)$. Define

$$B(p) = K_z(U_0(p), p).$$

Strict convexity implies that PC binds if and only if $B(p) \geq 0$. At any zero of B in $(0, p^A]$, using the left derivative at p^A,

$$B'(p) = K_{zp} + K_{zz}\big[u(i-d) - u(h)\big] < 0.$$

Hence B has at most one zero. At the endpoints,

$$\lim_{p\downarrow 0} B(p) = \frac{1}{u'(h)} > 0, \qquad B(p^A) = \frac{u'(h-d) - p^A[u'(i-d) - u'(h)]}{u'(h)u'(h-d)}.$$

If $p^A \leq \Gamma$, the boundary minimizes cost for every $p \leq p^A$. At p^A this is the zero contract, so proposition 1 implies $p^{HA} = p^A$. If $p^A > \Gamma$, there is a unique $\tilde{p} \in (0, p^A)$ with $B(\tilde{p}) = 0$. PC binds below this threshold and is slack above it. The negative derivative at p^A gives strictly positive profit there, so $p^{HA} > p^A$. For $p \in (p^A, p^{HA})$, PC is necessarily slack: positive profit implies $t < x$, and IC then gives wait utility strictly above $u(h-d) = U_0(p)$. Throughout the slack-PC region the unique optimum satisfies $K_z = 0$, and therefore

$$\frac{dU^{HA}(p)}{dp} = \frac{dz(p)}{dp} = -\frac{K_{zp}}{K_{zz}} > 0.$$

Finally, suppose A is nondecreasing. The function $s \mapsto u'(g(s))$ is concave, since its derivative is $-A(g(s))$. Binding IC and Jensen's inequality imply

$$u'(V) \geq pu'(I) + (1-p)u'(H) > p[u'(I) - u'(H)].$$

Thus eq. (4) cannot hold, and PC binds at every served type. As a positive-profit contract at $p \geq p^A$ necessarily has slack PC, we obtain $p^{HA} = p^A$. Applying the same inequality to autarky indifference yields $p^A < \Gamma$. □

Proof of Lemma 1. First consider a truthful menu. Type 1 necessarily visits, and truthfulness implies that every visit contract has the maximal net coverage k. Write

$$K = u(h-d+k), \qquad b = u(i-d+k) < K.$$

Assume for now that $\mathcal{W}$ is nonempty. Every wait contract has $u_i \leq b$ and $\Delta > 0$. A visit type cannot lie below a wait type: the latter's waiting payoff is strictly decreasing in type and is at least K at its assigned type. Thus $\mathcal{W}$ is an initial interval whenever it is nonempty. Truthfulness at type 0 bounds every healthy-state utility by $U(0)$, so

every wait type satisfies

$$K \le U(p) \le (1-p)U(0) + pb, \qquad p \le \frac{U(0)-K}{U(0)-b} < 1.$$

For $p < q$ in $\mathcal{W}$, the two reporting constraints give

$$(q-p)\Delta(q) \le U(p) - U(q) \le (q-p)\Delta(p). \tag{6}$$

Hence Δ is nonincreasing and bounded by the finite number $\Delta(0)$. It follows that U is Lipschitz on $\mathcal{W}$ and $\dot{U} = -\Delta$ almost everywhere. Comparison with types on the other side of the cutoff gives $U(p^*) = K$ whenever the cutoff contract is included, and $\lim_{p \uparrow p^*} U(p) = K$ otherwise. Therefore the extension $U = K$ above the cutoff is Lipschitz, nonincreasing, and convex. For a canonical menu, closedness of $\mathcal{W}$ now gives $\mathcal{W} = [0, p^*]$ and $\mathcal{V} = (p^*, 1]$.

Conversely, suppose conditions 1 and 2 hold. Monotonicity of Δ and the envelope formula imply

$$U(p) \ge U(q) + (q-p)\Delta(q) \qquad (p, q \in \mathcal{W}),$$

so no wait type gains by selecting another wait contract and waiting. Also $U(p) \ge U(p^*) = K$ on $\mathcal{W}$, and no contract offers a visit payoff above K. Let $L(p)$ be the waiting payoff under the common visit contract, and let $L_*(p) = U(p^*) + (p^* - p)\Delta(p^*)$. Since types immediately above p^* visit, $L(p^*) \le K = L_*(p^*)$. Maximality of k gives $u_i^V \ge u_i(p^*)$, so this cutoff inequality also implies $u_h^V \le u_h(p^*)$ (recall $p^* < 1$). The affine function $L_* - L$ is thus nonnegative at 0 and p^*, hence throughout $[0, p^*]$. Consequently, $L(p) \le L_*(p) \le U(p)$ on $\mathcal{W}$. Finally, every wait contract has $\Delta > 0$, so its waiting payoff at $p > p^*$ is no greater than at p^*, where it is at most K. All deviations are therefore unprofitable. □

Proof of Lemma 2. For $p < q$ in $\mathcal{W}$, truthfulness and monotonicity of Δ imply

$$u_h(p) - u_h(q) \ge p[\Delta(p) - \Delta(q)] \ge 0, \qquad u_i(q) - u_i(p) \ge (1-q)[\Delta(p) - \Delta(q)] \ge 0.$$

Thus, u_h is nonincreasing and u_i is nondecreasing on $\mathcal{W}$. Across the cutoff, maximality of k gives $u_i(p^*) \leq u_i(p^{*+})$, while the cutoff inequality established in the proof of Lemma 1 gives $u_h(p^*) \geq u_h(p^{*+})$. Both functions are constant on $\mathcal{V}$. The conclusions for t, $x - t$, and x follow from the utility–contract mapping. □

Proof of Lemma 3. For $p \in \mathcal{W}$, the envelope condition and spread bound give $U(p) \geq u(h) - p[u(h) - u(i-d)]$ and $u_i(p) = U(p) - (1-p)\Delta(p) \geq u(i-d)$. Hence $x(p) - t(p) \geq 0$. Since type p waits under his assigned contract, $U(p) \geq u(h - d + x(p) - t(p)) \geq u(h-d)$. Thus, $U(p)$ exceeds both autarky alternatives, so $U(p) \geq U_0(p)$. For $p \in \mathcal{V}$, $U(p) = U(p^*) \geq U_0(p^*) \geq U_0(p)$. □

Proof of Proposition 2. We first show that the principal's problem can be restricted without loss to a compact set of contracts. For a waiting type p, define

$$C_p(U, \Delta) \equiv p\, g\big(U - (1-p)\Delta\big) + (1-p) g\big(U + p\Delta\big).$$

This is the expected wealth cost of delivering expected utility U with utility spread Δ. Since $g = u^{-1}$ is increasing and convex, C_p is increasing in both arguments for $\Delta \geq 0$. Indeed,

$$\frac{\partial C_p}{\partial \Delta} = p(1-p)\left[g'\big(U + p\Delta\big) - g'\big(U - (1-p)\Delta\big)\right] \geq 0.$$

The principal's profit from a waiting type is $(1-p)h + p(i-d) - C_p(U, \Delta)$.

We claim first that it is without loss to impose $\Delta(p) \leq u(h) - u(i-d)$ on the wait region. Consider any feasible truthful canonical menu with nonnegative expected profit; menus with negative expected profit can be discarded because the zero contract is feasible and yields zero profit. Suppose first that $\Delta(p^*) \leq u(h) - u(i-d)$ but $\Delta(p) > u(h) - u(i-d)$ for some positive wait type. For every $p \in [0, p^*]$, define

$$\widehat{\Delta}(p) = \min\{\Delta(p), u(h) - u(i-d)\}, \qquad \widehat{U}(p) = U(p^*) + \int_p^{p^*} \widehat{\Delta}(s)\, ds,$$

and leave the visit contracts unchanged. The function $\widehat{\Delta}$ is nonincreasing, and the

modified allocation satisfies the envelope condition. Moreover, $\widehat{U}(p^*) = U(p^*)$ and $\widehat{\Delta}(p^*) = \Delta(p^*)$, so the terminal contract and terminal constraints are unchanged.

Let q be the supremum of the set of positive wait types at which $\Delta(p) > u(h) - u(i-d)$. By monotonicity, the original spread exceeds the bound on $(0, q)$ and satisfies it on $(q, p^*]$. Consequently, $\widehat{U}(p) = U(p)$ for $p \geq q$, whereas $\widehat{U}(p) = U(q) + (q-p)[u(h) - u(i-d)]$ for $p < q$. Participation is preserved because $U_0(p) - U_0(q) \leq (q-p)[u(h) - u(i-d)]$ and $U(q) \geq U_0(q)$. The reconstructed state-contingent utilities remain in the domain of g: $\widehat{u}_i(p) \geq u(i-d)$, $\widehat{u}_h(p) \leq u_h(0)$, and $\widehat{u}_i(p) \leq \widehat{u}_h(p)$. Furthermore, monotonicity of $\widehat{\Delta}$ and the envelope condition imply that $\widehat{u}_i$ is non-decreasing. Its terminal value is unchanged, so no modified wait contract offers net coverage above k. Together with the unchanged terminal and visit contracts, these properties preserve truthfulness by Lemma 1.

Finally, $\widehat{U}(p) \leq U(p)$ and $\widehat{\Delta}(p) \leq \Delta(p)$ for every wait type, with a strict decrease in the spread throughout $(0, q)$. Since C_p is increasing in U and strictly increasing in a positive spread for $p \in (0, 1)$, the expected wealth cost strictly falls on this interval and does not increase elsewhere. Full support of F therefore implies a strict increase in expected profit. Thus, every feasible menu in this case can be replaced by a strictly more profitable feasible menu satisfying the spread bound.

Suppose instead that

$$\Delta(p^*) > u(h) - u(i-d).$$

Monotonicity then implies $\Delta(p) > u(h) - u(i-d)$ for every $p \in (0, p^*]$. Individual rationality gives

$$U(p) \geq (1-p)u(h) + pu(i-d).$$

Hence, for every positive wait type,

$$C_p\big(U(p), \Delta(p)\big) > C_p\big((1-p)u(h) + pu(i-d), u(h) - u(i-d)\big) = (1-p)h + p(i-d),$$

so the principal earns strictly negative profit from that type. For a visit type, the

participation constraint gives $u(h-d+k) \geq u(h-d)$, hence $k \geq 0$, and profit is $-k \leq 0$. The menu is therefore weakly dominated by the zero contract. We may thus restrict attention, without changing the value of the problem, to menus satisfying

$$0 \leq \Delta(p) \leq u(h) - u(i-d).$$

We next show that it is without loss to impose $U(0) = u(h)$. Suppose $B \equiv U(0) - u(h) > 0$, and write $\Delta^* = \Delta(p^*)$ and $A = \int_0^{p^*} [\Delta(s) - \Delta^*] \, \mathrm{d}s$. For $\alpha \in [0,1]$, define

$$\widehat{\Delta}(p) = \Delta^* + (1-\alpha)[\Delta(p) - \Delta^*], \qquad \widehat{U}(p) = U(p^*) + \int_p^{p^*} \widehat{\Delta}(s) \, \mathrm{d}s$$

on the wait region, leaving the visit allocation unchanged. The terminal allocation, envelope condition, monotonicity, and spread bound are preserved, while $\widehat{U}(0) = U(0) - \alpha A$. Moreover, $\widehat{U} \leq U$ and $\widehat{\Delta} \leq \Delta$, so expected wealth cost weakly falls. For $j \in \{h, i\}$, $\widehat{u}_j(p) = (1-\alpha)u_j(p) + \alpha u_j(p^*)$, so the finite perturbation remains within the utility domain. Whenever $\widehat{U}(0) \geq u(h)$, the modified menu is feasible by Lemma 1 and Lemma 3.

If $0 < B \leq A$, choose $\alpha = B/A$. The modified menu satisfies $\widehat{U}(0) = u(h)$ and is strictly more profitable: $\alpha > 0$ and $A > 0$ imply that U or Δ strictly falls on a positive interval, which has positive probability by full support of F. If $B > A$, including $A = 0$, choose $\alpha = 1$. The modified wait contract is constant, with healthy-state utility $U(0) - A > u(h)$ and sick-state utility $U(0) - A - \Delta^* > u(i-d)$. Its premium is therefore negative and its net coverage positive, so $t - px = (1-p)t - p(x-t) < 0$ for every positive wait type. Visit types generate nonpositive profit. If the wait region contains positive types, full support therefore makes expected profit negative; if it contains only type 0, then $u(h-d+k) = U(0) > u(h)$, so $k > d$ and expected profit is again negative. Since the perturbation weakly raises profit, the original menu is also dominated by the zero contract. Thus the supremum is unchanged if we impose $U(0) = u(h)$ and $0 \leq \Delta \leq u(h) - u(i-d)$.

Normalize type 0's contract to $(0,0)$. This preserves his utility and profit and creates no profitable deviation because every other type already satisfies participation. For a positive wait type, the envelope condition, monotonicity, and spread bound imply $u_h(p) \leq u(h)$ and $u_i(p) \geq u(i-d)$. Hence $t(p) \geq 0$ and $x(p) - t(p) \geq 0$. Waiting also requires $x(p) < d$: otherwise visit wealth is weakly above healthy-state wealth and strictly above sick-state wealth. For visit types, participation and $U(p^*) \leq U(0) = u(h)$ imply $0 \leq k \leq d$. Replace their contracts by $(d-k, d)$. This preserves visit utility and profit and gives every type utility $u(h-d+k)$ from that report after choosing his action, so it creates no new profitable deviation. Consequently, all contracts can be restricted without loss to the compact set

$$\mathcal{C} = \{(t,x) \in \mathbb{R}^2 : 0 \leq t \leq x \leq d\}.$$

On the compact type and contract spaces $[0,1]$ and $\mathcal{C}$, the agent's utility is jointly continuous. The principal's payoff is jointly upper semicontinuous: at action indifference, the waiting payoff $t - px$ exceeds the visit payoff $t - x$ by $(1-p)x \geq 0$. The contract $(0,0)$ delivers U_0 to every type. Thus Theorem 4 of Backhoff-Veraguas et al. (2022), with a singleton prior and zero ambiguity penalty, supplies an optimal mechanism on $\mathcal{C}$. The compact reduction identifies its value with the unrestricted supremum, and canonicalization preserves expected profit. Lemma 2 makes the resulting contract schedules Borel measurable.

The strict improvements in the preceding capping and compression arguments now imply that every optimal canonical menu satisfies $U(0) = u(h)$ and $0 < \Delta(p) \leq u(h) - u(i-d)$ at every positive wait type. At type 0, the zero-contract normalization gives $\Delta(0) = u(h) - u(i-d)$.

Finally, the proof of lemma 1 gives $p^* < 1$. If $p^* = 0$, continuity and $U(0) = u(h)$ imply $u(h-d+k) = u(h)$, hence $k = d$. Because F has no atom at zero, expected profit would then be $-d < 0$, whereas the zero contract yields zero. Therefore $p^* \in (0,1)$. □

Proof of Theorem 2. By Proposition 2, $p^* \in (0,1)$ and $U(0) = u(h)$. Write $z(p) = x(p) - t(p)$ and $U^* = U(p^*)$. The optimality bounds and Lemma 2 imply $0 \leq t(p) \leq x(p) < d$ and $0 < \Delta(p) \leq u(h) - u(i-d)$ for positive wait types. Moreover, $U^* = u(h-d+k)$ and $k \geq z(p^*)$.

For any $q \in (0, p^*]$, consider the truncated menu M^q that keeps the contracts below q and assigns $(t(q), x(q))$ to every type $p \geq q$. We first verify feasibility. Let $W_r(p) = u_h(r) - p\Delta(r)$ denote type p's utility from waiting under the original contract of type r. For $r \leq q \leq p$,

$$W_q(p) - W_r(p) = W_q(q) - W_r(q) + (p-q)[\Delta(r) - \Delta(q)] \geq 0.$$

The first difference is nonnegative by truthfulness, and the second by monotonicity of Δ. Also, $z(q) \geq z(r)$, so visiting under q's contract is weakly preferred to visiting under any retained lower contract. Thus, every type above q optimally chooses q's contract, while lower types retain their original choices because the new menu is a subset of the old one. Participation is preserved: for $p \geq q$, the envelope condition and the spread bound give $W_q(p) \geq u(h) - p[u(h) - u(i-d)]$, while $z(q) \geq 0$ gives visit utility at least $u(h-d)$. Hence the utility from q's contract is at least $U_0(p)$.

Suppose first that $k > z(p^*)$ and apply this construction with $q = p^*$. Original wait types are unaffected. Every original visit type now generates profit at least $-z(p^*)$, since visiting gives $-z(p^*)$ and waiting gives $(1-p)x(p^*) - z(p^*) \geq -z(p^*)$. Previously each generated $-k$. Since $1 - F(p^*) > 0$, expected profit strictly increases, a contradiction. Therefore $k = z(p^*)$.

Suppose next that Δ is not constant on any left neighborhood of p^*. For $q < p^*$ sufficiently close to p^*, define $d_q = \Delta(q) - \Delta(p^*) > 0$ and $A_q = \int_q^{p^*} [\Delta(q) - \Delta(s)]\, ds$. Then $0 \leq A_q \leq (p^* - q)d_q$, and the envelope condition yields

$$W_q(p^*) = U^* - A_q, \qquad u_i(p^*) - u_i(q) = (1-p^*)d_q + A_q.$$

Since $k = z(p^*)$ and $g' = 1/(u' \circ g)$,

$$k - z(q) = g(u_i(p^*)) - g(u_i(q)) \geq \frac{(1-p^*)d_q}{u'(i-d)}.$$

Here both sick-state wealth levels lie in $[i-d, i]$, so $g' \geq 1/u'(i-d)$ on the relevant utility interval. Similarly, both visit wealth levels lie in $[h-d, h]$, and therefore

$$\begin{aligned} W_q(p^*) - u(h-d+z(q)) &= u(h-d+k) - u(h-d+z(q)) - A_q \\ &\geq \left[\frac{u'(h)(1-p^*)}{u'(i-d)} - (p^*-q)\right] d_q > 0 \end{aligned}$$

for q sufficiently close to p^*. Since W_q is decreasing, every type in $[q, p^*]$ continues to wait under M^q.

We now compare profits. For $p \in [q, p^*]$, truthfulness of the original menu gives $0 \leq u_h(q) - u_h(p) \leq p[\Delta(q) - \Delta(p)] \leq d_q$. Because healthy-state wealth is at most h, $t(p) - t(q) \leq d_q/u'(h)$. Moreover, $x(q) \leq x(p)$, so the loss from each such type is at most $d_q/u'(h)$. Every original visit type generates a gain of at least $k - z(q)$, regardless of his new action. Consequently,

$$\Pi(M^q) - \Pi(M) \geq d_q \left[\frac{(1-p^*)[1-F(p^*)]}{u'(i-d)} - \frac{F(p^*) - F(q)}{u'(h)}\right] > 0$$

for q sufficiently close to p^*, since F is continuous at p^* and $1 - F(p^*) > 0$. This contradicts optimality. Hence Δ is constant on some $[\bar{p}, p^*]$ with $\bar{p} < p^*$. The envelope condition then implies that u_h, u_i, t, and x are constant on this interval.

Finally, the terminal contract gives visit utility U^* and wait utility $U^* - (p - p^*)\Delta(p^*)$. Assigning it to all types above p^* therefore preserves their visit utility and the principal's profit, while inducing them to visit. This gives the outcome-equivalent representation. □

Proof of Proposition 3. By Proposition 2 and theorem 2, $p^* \in (0, 1)$ and the terminal constraint binds. Let $D = h - i > 0$, let (t^*, x^*) be the terminal wait contract, and write $k = x^* - t^*$, $V^* = h - d + k$, $I^* = V^* - D$, and $H^* = h - t^* = V^* + s^*$, where

$s^* = d - x^* > 0$. The optimality bounds give $k, t^* \geq 0$. Terminal indifference yields $u(V^*) = (1-p^*)u(H^*) + p^* u(I^*)$. Strict concavity implies $V^* < (1-p^*)H^* + p^* I^*$, and therefore

$$p^* < \frac{d - x^*}{D + d - x^*} \leq \frac{d}{D+d} = p^{SI}.$$

For part 1, define $\chi(V, s) = [u(V+s) - u(V)]/[u(V+s) - u(V-D)]$ for $s > 0$ and wealth levels in the domain of u. Then $p^* = \chi(V^*, s^*)$ and $p^A = \chi(h-d, d)$. Direct differentiation gives $\chi_s > 0$. Under nondecreasing absolute risk aversion, χ is nonincreasing in V. Indeed,

$$\frac{\chi(V,s)}{1-\chi(V,s)} = \frac{\int_0^s u'(V+y)\,\mathrm{d}y}{\int_{-D}^0 u'(V+y)\,\mathrm{d}y}.$$

For $V_2 > V_1$, the ratio $r(y) = u'(V_2+y)/u'(V_1+y)$ is nonincreasing because $(\log r)'(y) = A(V_1+y) - A(V_2+y) \leq 0$, where $A = -u''/u'$. Its weighted average, with weights $u'(V_1+y)$, on $[0,s]$ is therefore no greater than on $[-D, 0]$. This proves the claimed ordering of the odds and hence of χ.

Since $V^* \geq h - d$ and $s^* \leq d$,

$$p^* = \chi(V^*, s^*) \leq \chi(h-d, s^*) \leq \chi(h-d, d) = p^A.$$

If the allocation is not outcome-equivalent to no insurance, $x^* = k + t^* > 0$, so $s^* < d$ and the inequality is strict. Conversely, equality requires $s^* = d$, hence $k = t^* = 0$. Monotonicity and nonnegativity then imply zero contracts for positive wait types and zero net coverage for visit types, which is outcome-equivalent to no insurance. The no-insurance allocation has cutoff p^A.

For part 2, suppose that absolute risk aversion is nonincreasing and $p^A > \Gamma$. By Theorem 1, $p^{HA} > p^A$. Fix $p_0 \in (p^A, p^{HA})$ and a pointwise optimal contract (t_0, x_0) with $\pi_0 = t_0 - p_0 x_0 = \Pi^w(p_0) > 0$. This contract satisfies $0 < t_0 < x_0 < d$ and has binding action incentives at p_0. Consider the menu consisting of $(0,0)$ and (t_0, x_0). Every type in $[p^A, p_0]$ chooses the nonzero contract and waits: its wait utility is decreasing in p and at p_0 equals $u(h - d + x_0 - t_0) > u(h-d)$. Each such type

generates profit at least π_0. Types below p^A either choose the zero contract or choose the nonzero contract and wait, so their profit contribution is nonnegative. Types above p_0 generate profit greater than $-d$. Consequently, this menu yields expected profit at least

$$\pi_0[F(p_0) - F(p^A)] - d[1 - F(p_0)] \geq \pi_0(1 - 2\varepsilon) - d\varepsilon.$$

By contrast, any feasible menu with cutoff at most p^A earns nonpositive profit from visit types and at most d from each wait type. The latter bound follows from the observation in the proof of Proposition 1 that every positive-profit wait contract satisfies $0 < t < x < d$. Its expected profit is therefore at most $dF(p^A) \leq d\varepsilon$. Since $\varepsilon < \pi_0/[2(\pi_0 + d)]$, the first bound is strictly larger. Thus every optimal cutoff exceeds p^A.

To verify the sufficient condition in footnote 4, suppose an optimal CARA menu is nontrivial. Then $p^* \leq p^A$ and $k > 0$: if $k = 0$, the optimality bounds and monotonicity force $u_i = u(i - d)$ throughout the wait region, and truthfulness and $U(0) = u(h)$ then force zero wait contracts. Write $E = e^{\alpha(D+d)}$ and $z(p) = x(p) - t(p) \in [0, k]$. For a wait type, PC gives

$$(1 - p)(e^{\alpha t(p)} - 1) \leq pE(1 - e^{-\alpha z(p)}).$$

Using $e^a - 1 \geq a$ and $1 - e^{-a} \leq a$, its profit satisfies

$$t(p) - px(p) \leq p(E - 1)z(p) \leq p^A(E - 1)k = (e^{\alpha d} - 1)k.$$

The final bound is strict for every positive wait type: if $z(p) > 0$ the exponential inequality is strict, while if $z(p) = 0$ its profit is nonpositive. Visit types yield $-k$. Since $F(p^*) > 0$,

$$\Pi < k\{e^{\alpha d}F(p^*) - 1\} \leq k\{e^{\alpha d}F(p^A) - 1\} \leq 0,$$

contradicting the availability of the zero contract. □

Proof of Proposition 4. The zero contract is feasible and gives zero profit, so $P(u_\alpha, w, D) \geq 0$. Consider any truthful menu and any type p. The contract and action assigned to this type are feasible in the pointwise hidden-action problem in which p is observable. Hence the profit generated by type p cannot exceed the value of that pointwise problem.

Under CARA, Theorem 1 gives $p^{HA} = p^A$, where

$$p^A = \frac{e^{\alpha d} - 1}{e^{\alpha(D+d)} - 1}.$$

No type $p \geq p^A$ can generate positive profit. If $p < p^A$ and an assigned wait-inducing contract generates positive profit, the observation in the proof of Proposition 1 implies $0 < t < x < d$, and therefore $t - px < t < d$. A visit-inducing contract gives nonpositive profit. Integrating these pointwise bounds yields

$$P(u_\alpha, w, D) \leq d\, F(p^A).$$

For fixed $D, d > 0$, $p^A \to 0$ as $\alpha \to \infty$, while for fixed $\alpha, d > 0$, $p^A \to 0$ as $D \to \infty$. Since F is continuous at zero, $F(p^A) \to 0$ in both cases. The two conclusions follow from the preceding upper bound. Positive profit is possible at finite CARA parameters: the construction in the proof of proposition 1 gives a contract with strict participation, strict waiting incentives, and uniformly positive profit on a neighborhood of any fixed $p_b \in (0, p^A)$. Offering it together with $(0, 0)$ yields positive expected profit for a smooth full-support density sufficiently concentrated on that neighborhood, since losses elsewhere are bounded by d. Holding this distribution fixed, either limit above therefore implies a strict profit decline. □

Verification of Example 1. Let $\widehat{p} = 2/5$. Under log utility and $D = d = 1$,

$$p^A(w) = \frac{\log w - \log(w-1)}{\log w - \log(w-2)} \qquad \text{and} \qquad \Gamma(w) = \frac{w(w-2)}{2(w-1)}.$$

Thus, $p^A(4) \simeq 0.415 > \widehat{p} > p^A(3) \simeq 0.369$, while $\Gamma(4) = 4/3$ and $\Gamma(3) = 3/4$. Hence $p^A(w) \leq \Gamma(w)$ at both wealth levels, and Theorem 1 gives $p^{HA}(w) = p^A(w)$. The

pointwise hidden-action problem therefore yields strictly positive profit from type $\widehat{p}$ when $w = 4$. When $w = 3$, every wait-inducing contract for this type yields negative profit, while a visit-inducing contract yields nonpositive profit; since the zero contract is available, the overall pointwise value is zero.

Choose an open interval N containing $\widehat{p}$ such that every $p \in N$ lies below $p^{HA}(4)$ and above $p^{HA}(3)$. In the environment with $w = 4$, choose a contract (t, x) that strictly satisfies participation and incentive compatibility at $\widehat{p}$ and generates positive profit. Such a contract exists by the construction in the proof of Proposition 1. By continuity, after shrinking N if necessary, there exists $\eta > 0$ such that every $p \in N$ accepts (t, x), waits, and generates profit at least η. We may choose the contract so that $0 < t < x < d$. In the environment with $w = 3$, the overall pointwise value of every type in N is zero.

Let ψ_N be a smooth density supported in N and, for $\varepsilon \in (0, 1)$, define

$$f_\varepsilon(p) = (1 - \varepsilon)\psi_N(p) + \varepsilon, \qquad p \in [0, 1].$$

This density is smooth and strictly positive on $[0, 1]$, and assigns probability at least $1 - \varepsilon$ to N. When $w = 4$, offering the zero contract and (t, x) gives expected profit at least $(1 - \varepsilon)\eta - \varepsilon d$: types in N generate at least η, while profit from every type outside N is bounded below by $-d$. When $w = 3$, types in N generate nonpositive profit, while positive profit from every type outside N is bounded above by d. Hence, for the distribution with density f_ε,

$$P(u, 4, 1) \geq (1 - \varepsilon)\eta - \varepsilon d, \qquad P(u, 3, 1) \leq \varepsilon d.$$

Choosing $\varepsilon < \eta/(\eta + 2d)$ gives $P(u, 4, 1) > P(u, 3, 1)$. □

Proof of Proposition 5. We first show that the contract schedule is continuous. Since Δ is nonincreasing, any interior discontinuity must be a downward jump. Sup-

pose that, for some $q \in (0, p^*)$,

$$\Delta^- \equiv \Delta(q^-) > \Delta(q^+) \equiv \Delta^+.$$

For $\varepsilon > 0$ sufficiently small, let $a_\varepsilon = q - \varepsilon$ and $b_\varepsilon = q + \varepsilon$, and replace U on $[a_\varepsilon, b_\varepsilon]$ by the chord joining its endpoint values:

$$\widehat{U}_\varepsilon(p) = \frac{b_\varepsilon - p}{2\varepsilon} U(a_\varepsilon) + \frac{p - a_\varepsilon}{2\varepsilon} U(b_\varepsilon).$$

Leave U unchanged outside this interval. The associated utility spread is constant on $(a_\varepsilon, b_\varepsilon)$ and equal to

$$\overline{\Delta}_\varepsilon \equiv \frac{U(a_\varepsilon) - U(b_\varepsilon)}{2\varepsilon}.$$

Because U is convex, its chord lies weakly above it. Moreover, since Δ is nonincreasing, $\overline{\Delta}_\varepsilon$ lies between the spreads immediately to the left and right of the interval. Hence the modified schedule remains convex, satisfies $\widehat{U}'_\varepsilon = -\widehat{\Delta}_\varepsilon$, and preserves $0 \leq \widehat{\Delta}_\varepsilon \leq u(h) - u(i - d)$. It leaves $U(0)$, the terminal contract, and the terminal net-coverage constraint unchanged. Since $\widehat{U}_\varepsilon \geq U$, the participation constraint is preserved, and Lemmas 1 and 3 imply that the modified menu is feasible.

For a waiting type, let

$$C(p, U, \Delta) \equiv p\, g\big(U - (1-p)\Delta\big) + (1-p) g\big(U + p\Delta\big)$$

be the expected wealth cost, where $g = u^{-1}$. For $p \in (0, 1)$,

$$C_{\Delta\Delta}(p, U, \Delta) = p(1-p)\left[(1-p) g''\big(U - (1-p)\Delta\big) + p\, g''\big(U + p\Delta\big)\right] > 0,$$

so C is strictly convex in Δ. Since $\overline{\Delta}_\varepsilon \to (\Delta^- + \Delta^+)/2$, a change of variables $p = q + \varepsilon y$ and dominated convergence give

$$\begin{aligned}\lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon} \int_{a_\varepsilon}^{b_\varepsilon} f(p) \Big[C(p, \widehat{U}_\varepsilon(p), \overline{\Delta}_\varepsilon) - C(p, U(p), \Delta(p))\Big]\, \mathrm{d}p \\ = f(q) \left[2C\left(q, U(q), \frac{\Delta^- + \Delta^+}{2}\right) - C(q, U(q), \Delta^-) - C(q, U(q), \Delta^+)\right] < 0.\end{aligned}$$

The modified menu therefore has strictly lower expected wealth cost for ε sufficiently small, contradicting optimality. Thus, Δ is continuous on $(0, p^*)$. Since U is continuous and $u_h(p) = U(p) + p\Delta(p)$ and $u_i(p) = U(p) - (1-p)\Delta(p)$, the functions u_h, u_i, t, and x are continuous as well.

We next derive the local curvature condition. On a separating interval, $U' = -\Delta$ gives

$$u_i'(p) = -(1-p)\Delta'(p), \qquad u_h'(p) = p\Delta'(p).$$

It follows that

$$t'(p) = -p\, g'(u_h(p))\Delta'(p)$$

and

$$x'(p) = -\left[(1-p)g'(u_i(p)) + p\, g'(u_h(p))\right]\Delta'(p).$$

Both derivatives are strictly positive, and therefore

$$T'(x(p)) = \frac{t'(p)}{x'(p)} = \frac{p\, u'(I(p))}{(1-p)u'(H(p)) + p\, u'(I(p))}.$$

Since $\Delta(p) > 0$, we have $I(p) < H(p)$ and $u'(I(p)) > u'(H(p))$, which yields $p < T'(x(p)) < 1$.

On a separating interval, the monotonicity and spread bounds are locally slack. The Euler equation for minimizing expected wealth cost, subject to $U' = -\Delta$, is

$$(fC_\Delta)' = -fC_U.$$

Write

$$a_i = g'(u_i), \quad a_h = g'(u_h), \quad b_i = g''(u_i), \quad b_h = g''(u_h), \quad r = p(1-p),$$

and define

$$B = (1-p)b_i + pb_h, \qquad S = (1-p)\frac{b_i}{a_i} + p\frac{b_h}{a_h}, \qquad c = \frac{pb_h/a_h}{S} \in (0,1), \qquad \kappa = \frac{(a_h - a_i)S}{B} > 0.$$

Define $m(p) = pa_h/[(1-p)a_i + pa_h]$ on the whole wait interval; it equals $T'(x(p))$ at

separating points. Since $C_\Delta = r(a_h - a_i)$, $C_U = pa_i + (1-p)a_h$, and $m/(1-m) = pa_h/[(1-p)a_i]$, expanding the Euler equation and differentiating the latter identity gives

$$\frac{m'}{m(1-m)} = \kappa \left[\frac{3p-2+c}{p(1-p)} - \frac{f'}{f} \right]. \tag{7}$$

Here

$$c = \frac{pA(H)/u'(H)}{(1-p)A(I)/u'(I) + pA(H)/u'(H)}.$$

Because $x' > 0$, this identity determines the sign of T''.

The Euler equation solves for $\Delta' = Q(p, U, \Delta)$ with Q continuous. At a partition boundary p_0 with $\Delta'(p_0) < 0$, the mean value theorem gives separating points approaching from each side, hence $Q(p_0) = \Delta'(p_0) < 0$. Near p_0, Q stays negative. On each adjacent C^1 piece, the continuous derivative Δ' is either zero or Q, so it cannot switch between them. The approaching separating points select Q on both sides. Thus eq. (7) extends to p_0, and $x'(p_0) > 0$ ensures that the corresponding tariff curvature exists.

We next establish the global curvature ordering. Write

$$L = \frac{f'}{f}, \qquad G = \frac{3p-2+c}{p(1-p)}, \qquad \Psi = L - G.$$

Let $R = A(I)/A(H)$. Since I is nondecreasing, H is nonincreasing, and A is nonincreasing, R is nonincreasing. Moreover,

$$\frac{c}{1-c} = \frac{1}{R}\frac{m}{1-m}, \qquad \frac{c'}{c(1-c)} = \frac{m'}{m(1-m)} - \frac{R'}{R}.$$

Consequently $c' \geq -c(1-c)\kappa\Psi_+$ at separating points, where $\Psi_+ = \max\{\Psi, 0\}$. At a differentiability point with $\Delta' = 0$, both state-wealth derivatives vanish and $c' = c(1-c)/r > 0$, so the same bound holds there. Thus, almost everywhere,

$$G' \geq \beta - K\Psi_+, \qquad \beta = \frac{3p^2 - 4p + 2 - c(1-2p)}{r^2} > 0, \qquad K = \frac{c(1-c)\kappa}{r}.$$

The numerator defining β equals

$$(1-c)(3p^2-4p+2)+c(3p^2-2p+1)>0.$$

Continuity, monotonicity, and the finite piecewise-C^1 assumption make Δ, and hence G, locally absolutely continuous. On every compact subinterval of $(0,p^*)$, therefore, $\beta\geq\beta_0>0$ and $K\leq K_0<\infty$. Suppose $\Psi(s)\leq 0<\Psi(t)$ for some $s<t$ in this subinterval. Choose $0<\eta<\Psi(t)$ with $K_0\eta<\beta_0$. Let v be the first point after s at which $\Psi(v)=\eta$, and let u be its last preceding zero. On (u,v), $0<\Psi<\eta$, so $G'>0$ almost everywhere and $G(v)>G(u)$. Log-concavity makes L nonincreasing, giving

$$\Psi(v)=L(v)-G(v)<L(u)-G(u)=0,$$

a contradiction. Once Ψ is nonpositive it therefore stays nonpositive, and the same differential bound then makes it strictly negative at every later point. Hence Ψ has at most one zero, with positive values before it and negative values after it, allowing either sign region to be empty. Equation (7) proves the asserted curvature ordering.

Finally, let $[p_1,p_2]\subset(0,p^*)$ be a pooling interval bordered by separating regions, with common contract (t_0,x_0) and state wealths $I_0<H_0$. Continuity and the slope formula give

$$T'_-(x_0)=m_0(p_1),\qquad T'_+(x_0)=m_0(p_2),\qquad m_0(p)=\frac{pu'(I_0)}{(1-p)u'(H_0)+pu'(I_0)}.$$

Since

$$m_0'(p)=\frac{u'(I_0)u'(H_0)}{[(1-p)u'(H_0)+pu'(I_0)]^2}>0,$$

the pool produces a strict upward kink. □

References

Backhoff-Veraguas, J., Beissner, P., and Horst, U. (2022). Robust contracting in general contract spaces. *Economic Theory*, 73(4):917–945.

Boone, J. (2015). Basic versus supplementary health insurance: moral hazard and adverse selection. *Journal of Public Economics*, 128:50–58.

Castro-Pires, H., Chade, H., and Swinkels, J. (2024). Disentangling moral hazard and adverse selection. *American Economic Review*, 114(1):1–37.

Chade, H., Marone, V. R., Starc, A., and Swinkels, J. (2022). Multidimensional screening and menu design in health insurance markets. Technical report, National Bureau of Economic Research.

Chade, H. and Schlee, E. (2012). Optimal insurance with adverse selection. *Theoretical Economics*, 7(3):571–607.

Finkelstein, A., Taubman, S., Wright, B., Bernstein, M., Gruber, J., Newhouse, J. P., Allen, H., Baicker, K., and Group, O. H. S. (2012). The oregon health insurance experiment: evidence from the first year. *The Quarterly journal of economics*, 127(3):1057–1106.

Jullien, B., Salanie, B., and Salanie, F. (1999). Should more risk-averse agents exert more effort? *The Geneva Papers on Risk and Insurance Theory*, 24:19–28.

Jullien, B., Salanie, B., and Salanie, F. (2007). Screening risk-averse agents under moral hazard: single-crossing and the cara case. *Economic Theory*, 30:151–169.

Newhouse, J. P. (1993). *Free for all?: lessons from the RAND health insurance experiment*. Harvard University Press.

Pauly, M. V. (1968). The economics of moral hazard: comment. *The american economic review*, 58(3):531–537.

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.