SIEPR Discussion Paper No. 17-007

Subrogation and the Theory of Insurance When Suits Can Be Brought for Losses Suffered

By

A. Mitchell Polinsky and Steven Shavell

Stanford Institute for Economic Policy Research
Stanford University
Stanford, CA 94305
(650) 725-1874

March 2017

Subrogation and the Theory of Insurance

When Suits Can Be Brought for Losses Suffered

A. Mitchell Polinsky and Steven Shavell*

Abstract: The theory of insurance is considered here when an insured individual may be able to sue another party for the losses that the insured suffered—and thus when an insured has a potential source of compensation in addition to insurance coverage. Insurance policies reflect this possibility through so-called subrogation provisions that give insurers the right to step into the shoes of insureds and to bring suits against injurers. We show that subrogation provisions are a fundamental feature of optimal insurance contracts because they relieve litigation-related risks and result in lower premiums—financed by the litigation income of insurers. This income includes earnings from suits that insureds would not otherwise have brought. We also characterize optimal subrogation provisions in the presence of loading costs, moral hazard, and non-monetary losses.

Key words: insurance; subrogation; accidents; torts; litigation; loading costs; moral hazard; non-monetary losses

---

* Josephine Scott Crocker Professor of Law and Economics at Stanford Law School and Research Associate, National Bureau of Economic Research; and Samuel R. Rosenthal Professor of Law and Economics at Harvard Law School and Research Associate, National Bureau of Economic Research, respectively. We thank Kenneth Abraham, Scott Baker, Tom Baker, Mark Geistfeld, Daniel Gottlieb, Oliver Hart, Louis Kaplow, Saul Levmore, Stephen Mildenhall, Ariel Porat, David Rosenberg, Ted Sichelman, Kathryn Spier, and participants at seminars at Harvard, the NBER, and the Sixth Annual Law and Economic Theory conference at NYU for comments; James Fox, Anastasios Moraitis, and Paul Riskind for research assistance; and the John M. Olin Program in Law and Economics at Stanford Law School and the John M. Olin Center for Law, Economics, and Business at Harvard Law School for research support. [Email contacts: <polinsky@stanford.edu> and <shavell@law.harvard.edu>.]

1. Introduction

In this article we consider the theory of insurance when insured individuals may be able to sue other parties for the losses they have suffered—and thus have a potential source of compensation in addition to insurance. Such situations are ubiquitous: an insured driver whose car is damaged in an accident may be able to sue another driver who caused the accident for repair costs; an insured homeowner whose house burns down due to a gas leak may be able to sue the utility company for the value of his residence; or a person covered by health insurance who slips and falls in a store may be able to sue the proprietor for medical expenses.

Insurance policies reflect the reality that it is often possible for an insured to sue an injurer for losses sustained. Namely, insurance policies not only promise to compensate insureds for their losses, they also usually include what are known as subrogation provisions that accord the insurer the right to step into the shoes of an insured and to sue a party who caused the insured's losses.[1] Suppose that driver A owns collision insurance coverage on his car and that driver B negligently causes an accident that damages A's car. Driver A would receive payment from his insurer for repair costs after the accident and then, under the subrogation terms of his insurance policy, driver A's insurer could sue driver B. Such a subrogated insurer typically would be authorized to retain much or all of the proceeds from suit[2] and also would bear much or all of the litigation costs.[3]

[1] For a general description of subrogation, see Jerry and Richmond (2012, pp. 648-54). See also Abraham and Schwarcz (2015, pp. 266-68), Baker (2008, pp. 331-33), and Harrington and Niehaus (2004, p. 195). Subrogation provisions may be explicit in insurance contracts, supplied by statute, or legally presumed to apply; see Jerry and Richmond (2012, pp. 651–54). In some instances, subrogation effectively comes about when insureds themselves bring suit and then reimburse insurers for the coverage that they had received; see Couch on Insurance (2015, §226:1, 3, 4, 41). On the origins of subrogation, which date from antiquity, see Marasinghe (1976a, 1976b).

[2] On the practices and legal doctrines governing the amount that subrogated insurers retain from judgments, see generally Couch on Insurance (2015, §223). The subrogation income that insurers collect can be significant; for

In the United States, subrogation provisions are a common feature of property insurance, liability insurance, health and medical insurance, and disability insurance policies.[4] Subrogation provisions may also apply to governmentally-provided insurance, notably, to workers compensation, Medicare, and Medicaid programs.[5] The prevalence of subrogation provisions is similar in other countries as well.[6]

Our object here is to demonstrate that subrogation provisions are a fundamental feature of optimal insurance contracts—those that maximize the expected utility of insureds subject to the constraint that the insurer covers its expected expenses—and also to study the specific character of optimal subrogation provisions.[7]

To this end, we analyze a model in which a risk-averse individual who confronts the risk of an accident may purchase insurance and, if an accident occurs, may (or his subrogated insurer may) sue the injurer. Suit is assumed to be costly and to be successful only with a probability, in which event the injurer must pay damages equal to harm.

In Section 2, we show in a basic version of the model that insurance policies with subrogation provisions are superior to pure insurance policies—policies that pay coverage after

---

example, a study cited in Insurance Day (2015) found that for automobile insurers such income amounted to 27% of their paid losses for collisions.

[3] The insurer would usually be responsible for the entirety of litigation costs when the insurer retains the full judgment but only for a proportional share of the costs when the insurer keeps less than the full judgment; see Couch on Insurance (2015, §223: 113, 119).

[4] See, for example, Dobbyn and French (2016, pp. 425-28) on property, liability, health and medical insurance, and Dobbyn and French (2016, p. 46) and Bleed (2001, p. 734) on disability insurance.

[5] See, for example, Dobbyn and French (2016, pp. 429-30) on workers compensation and Jerry and Richmond (2012, p. 652) on Medicare and Medicaid.

[6] See, for example, Insurance Day (2009), reviewing subrogation worldwide, including in England, Germany, France, Russia, Spain, Switzerland, Brazil, India, Singapore, China, and Australia.

[7] As we note in Section 4, however, we do not consider the effects of subrogation on social welfare, through its influence on the costs of litigation and the deterrence of harm.

an accident occurs but that do not address the possibility of suit. Under a pure insurance policy, we find that optimal coverage would generally be different from full coverage because the insured might bring and win a suit after an accident occurred. Notably, the insured would tend to purchase less than full coverage because the possibility of receiving damages from a successful suit would reduce his need for compensation.[8] That an insured's coverage under a pure insurance policy would differ from full coverage would, however, leave him exposed to risk from accidents, and litigation itself would also entail risk. Moreover, because of litigation risk, the insured would refrain from pursuing a range of positive expected value suits.

The foregoing problems of risk-bearing and of the failure to bring all positive expected value suits can be eliminated by including a subrogation provision in an insurance policy.[9] Under such a policy, suppose that insurance coverage is full, that the insurer would sue whenever a suit has positive expected value, and that the insurer would retain the entirety of damages from a successful suit. Then insureds would be better off than under the optimal pure insurance policy: they would not bear risk from accidents; they would not bear risk from litigation; and they would benefit from all positive expected value suits—through a reduction in their premiums that reflects the income that insurers would obtain from suit.

In Section 3, we study the nature of optimal insurance policies with subrogation provisions under three natural variations of our initial assumptions. We first assume that insurers bear positive administrative costs (such as the expense of checking the veracity of claims) that rise with the level of coverage. Hence, insurance premiums will include a loading above their

---

[8] Although optimal coverage would tend to be less than full coverage for this reason, it could in principle be anywhere in the interval between the litigation cost and the loss plus the litigation cost; see Proposition 1(b).

[9] The formal definition of a subrogation provision involves instructions to the insurer when to pursue suits and the share of damages to be returned to the insured if suit is successful.

actuarially fair level, and optimal insurance coverage will be less than full for well-recognized reasons. We show that under the optimal subrogation provision, the shortfall of optimal insurance coverage from the loss due to an accident will be alleviated by the insured's receiving a positive share of damages from a successful suit—in other words, the subrogated insurer will not retain all of the damages. This will be desirable because the receipt of a share of damages by the insured constitutes a form of supplemental insurance coverage that is effectively free of administrative costs.[10] For a similar reason, it will be in the insured's interest for some negative expected value suits to be pursued by the insurer.

We next consider moral hazard. We assume that insureds can reduce the probability of one kind of accident by the exercise of unobservable care (for example, homeowners can reduce the risk of fire by storing flammables away from heat sources), whereas insureds play no role in the occurrence of a different kind of accident that is caused by others (home fires can arise due to the negligence of contractors). We verify that optimal insurance coverage will be less than complete for the familiar reason that some exposure to risk will provide an incentive for insureds to take care to reduce accident risk—here the risk of the first kind of accident. We then prove that, because optimal insurance coverage is less than full, the optimal subrogation provision will award the insured a positive share of the damages from a successful suit. The logic is two-fold. On one hand, the receipt of a share of damages will help to offset the portion of the loss not covered by insurance (as was true when insurance coverage is less than full due to administrative costs). On the other hand, the receipt of a share of damages will not contribute to moral hazard. That is because of the presumption that for a suit to be successful, the accident must be found by

[10] The administrative costs at issue are negligible, as they are only those associated with transferring funds to the insured. (This statement is not inconsistent with the fact that litigation costs are distinctly positive.) See Section 3.A.

a court to have been caused by another party (a contractor rather than a homeowner). We also show that it is desirable for the insurer to pursue some negative expected value suits.

Last we examine accidents that include a non-monetary utility loss, such as that due to the death of a spouse. In these circumstances, it is well-understood that optimal insurance coverage would not reflect the non-monetary loss, presuming that it does not affect the marginal utility of wealth; optimal coverage would be restricted to the monetary loss (associated with the income that the spouse would have earned). Courts, however, award damages for both types of losses. It follows that under the optimal subrogation provision, the insurer would retain all of the damages from suit, including the component of damages for the non-monetary loss. The insured would of course benefit from the insurer's receipt of damages for the non-monetary loss because that would contribute to the reduction in his insurance premium. Also, under the optimal subrogation provision, the insurer would bring suit if and only if the expected return is positive.

We conclude in Section 4 with a number of observations about factors that are not included in our model, about why the sale of claims, which might be thought to be a substitute for subrogation, does not occur in practice, and about various welfare-reducing legal restrictions on subrogation.

Before proceeding, we note that we are not aware of any prior articles in the economic literature on insurance that study subrogation,[11] and of only very limited writing on this topic in law and economics literature.[12]

[11] For example, none of the thirty-seven articles in Dionne (2013)—*Handbook of Insurance*—considers subrogation.

[12] Shavell (1987, p. 255-56) shows that subrogation provisions are desirable for insureds, but in a model with costless litigation and thus in which no decisions are made about the bringing of suit; see also note 26 below. Sykes (2001) considers optimal subrogation provisions in the presence of loading, moral hazard, and non-monetary losses, but does not analyze the basic rationale for subrogation. In his analysis, he makes assumptions that in our opinion lack economic justification and that generate conclusions that we believe are incorrect in all three settings that he studies; see notes 31, 37, and 40 below. Additionally, Gomez and Penalva (2015) focus on the effects of

2. Basic Analysis

We consider here the standard model of insurance for accident risks, but supplemented by the ability of insured individuals to bring suit for harm that they suffer. In particular, we assume that identical risk-averse individuals face a chance of suffering harm from an accident; that they purchase insurance policies that maximize their expected utility; and that if an accident occurs, they will be able to sue a potentially liable party for recompense, where suit would involve a litigation cost and result in success only with a probability.[13]

Specifically, define the following notation.

y = initial wealth of an individual;

$U(\cdot)$ = utility of an individual from wealth, where $U' > 0$ and $U'' < 0$;

p = probability of an accident; $p \in (0, 1)$;

h = harm if an accident occurs; $h > 0$;

k = cost of bringing a suit; $k \geq 0$; and

q = probability of winning a suit, resulting in an award of h; $q \in [0, 1]$.

For simplicity, we assume that there is a single value of each of the above variables (and thus that these values are known to all parties).

A *pure insurance policy* is defined by a premium that the insured pays at the outset and a coverage amount that the insured receives immediately after an accident occurs but before a suit

---

subrogation on the deterrence of accidents, not on contractually optimal subrogation provisions; see note 54 below. Finally, Viscusi (1989) investigates empirically the role of subrogation in product liability suits brought by workers compensation systems on behalf of injured employees. On legal literature concerned with subrogation (which often reflects economic views), see, for example, Abraham (1986, pp. 153-55), Kimball and Davis (1962), and Reinker and Rosenberg (2007), as well as articles cited in Couch on Insurance (2015, Part XI).

[13] We abstract from the possibility of settlement in the model for the following reason. If settlement were allowed, all suits would settle (since, as will be seen, information about suit is symmetric). Consequently, one of the primary benefits of a subrogation policy—that it can reduce or eliminate the risk of litigation—would not be captured. (Had we employed a model in which both suit and settlement can occur, due to asymmetric information, some of the analysis in section 3 would have become unnecessarily complicated.)

might be brought by the insured. This assumption about the time at which coverage is received reflects two aspects of reality: that a person will frequently have consumption needs that he must satisfy shortly after an accident (like replacing a car that he requires for transportation); and that litigation outcomes (or settlements) often take a substantial time to eventuate.[14]

An *insurance policy with a subrogation provision* adds to the pure insurance policy an instruction to the insurer whether to bring a suit on behalf of the insured[15] after an accident occurs,[16] states that the insurer will bear litigation costs k if a suit is brought,[17] and specifies the share (perhaps zero) that the insurer will pay to the insured from the award h if a suit is successful.

Let

π = insurance premium;

c = insurance coverage; $c \geq 0$;

ϕ = instruction whether the insurer sues under a subrogation provision; if $\phi = 0$, suit is not brought; if $\phi = 1$, suit is brought; and

s = insured's share of the award h under a subrogation provision if a suit is successful; $s \in [0, h]$.

---

[14] In the course of analyzing the model presented here, we investigated a model in which consumption occurs at two times—directly following an accident and after the outcome of litigation. In the latter framework, we demonstrated that it was optimal for the insured to receive compensation immediately following an accident. This two-consumption-period model did not provide much understanding beyond that gained from the model that we have described in which insurance payments are assumed to be made just after an accident.

[15] We assume that the insured may not bring suit himself if the policy has a subrogation provision.

[16] In practice, we would not expect insurance contracts to have explicit instructions about when the insurer should bring a suit due to problems of asymmetry of information; see Section 4.

[17] This assumption is made for simplicity; it would be straightforward to show that it is a feature of the optimal contract.

Thus, a pure insurance policy is described by a premium π and a coverage level c; and a policy with a subrogation provision is described by a premium π, a coverage level c, an instruction ϕ about suit, and the insured's share s of an award.

We will determine the optimal pure insurance policy and the optimal insurance policy with a subrogation provision, and then will compare the two. By the optimal policy, we mean the policy (among either pure policies or policies with a subrogation provision, as the case may be) that maximizes the expected utility of the insured subject to the constraint that the insurer breaks even—that the premium paid equals the expected expenses of the insurer.[18]

Under a pure insurance policy, the premium constraint is

$$\pi = pc. \tag{1}$$

The expected utility of the insured given c will depend on whether he would bring a suit if an accident occurred. As is evident from Figure 1,[19] if the insured would not bring a suit, then his expected utility would simply be

$$EU_N(c) = (1 - p)U(y - \pi) + pU(y - \pi - h + c). \tag{2}$$

If the insured would bring a suit, his expected utility would be

$$EU_S(c) = (1 - p)U(y - \pi) + p[qU(y - \pi + c - k) + (1 - q)U(y - \pi - h + c - k)], \tag{3}$$

where the harm h does not appear in the first term in brackets because the insured receives h from a successful suit. [Insert Figure 1 here.] Therefore, given any c, the insured will bring a suit

[18] The justification for relying on the expected value of the insurer's expenses is the standard one in the insurance literature, namely, that the insured risks are independent and that the law of large numbers implies that if the premium is above the expected expenses by any small ε, then the probability that the actual expenses would exceed the sum of premiums tends to zero as the number of insureds increases.

[19] We are implicitly assuming that the insured cannot sell his claim to the insurer or to another party. We comment on this possibility in Section 4.

after an accident occurs if and only if (3) exceeds (2), and his expected utility will be $max[EU_N(c), EU_S(c)]$.[20] Consequently, the optimal pure insurance policy solves the problem

(4) $$max\{max[EU_N(c), EU_S(c)]\} \text{over } c.$$

The solution to (4) is described in the next result, where c^* denotes the optimal level of coverage.

Proposition 1. Under the optimal pure insurance policy,

(a) if suits are not optimal to bring, optimal coverage is full, $c^* = h$, and each insured thus has certain wealth of $y - ph$;

(b) if suits are optimal to bring, optimal coverage c^* is in $(k, h + k)$, where c^* is determined by

(5) $$U'(y - pc) = qU'(y - pc + c - k) + (1 - q)U'(y - pc - h + c - k),$$

and each insured has expected utility of $EU_S = (1 - p)U(y - pc^*) + p[q\text{U}(y - pc^* + c^* - k) + (1 - q)U(y - pc^* - h + c^* - k)]$; and

(c) suits are optimal to bring if and only if their expected value exceeds a positive threshold: given any q and h, there exists a t satisfying $0 < t < qh$ such that if $qh - k > t$, suit is optimal to bring; and if $qh - k \leq t$, suit is not optimal to bring.

Notes. (i) In part (a) the problem for the insured is the standard insurance problem, in which, as is well known, optimal coverage is full.[21]

(ii) The intuition underlying (b) is that, if an accident occurs, the insured will bear k in litigation costs, suggesting that he would want to insure for at least that amount; and since he will also bear the risk of losing the suit, he might want to insure for as much as h more. It is plausible,

---

[20] We assume for concreteness that if the insured is indifferent between suing and not suing, he will not sue, and we will adopt similar conventions below without comment.

[21] See, for example, Schlesinger (2013, p. 170).

however, that c^* would be significantly less than $h + k$, and, indeed, less than h. The reason is that c would often result in excessive total compensation, for if the suit is successful, the insured collects damages of h in addition to c. This point is reinforced by the fact that the probability q that suit would be won will tend to be high, since the hypothesis is that suit is optimal to bring.

(iii) To illustrate (b), suppose that the utility function is $U(y) = -1/y$, and let y = \$200,000, $p = 0.3$, h = \$100,000, and k = \$15,000.[22] If q, the probability of winning at trial, is 0.99, c^* = \$17,993,[23] not much more than litigation costs and far below the accident loss. If $q = 0.8$, c^* = \$53,365; in effect, insurance then covers the \$15,000 of litigation costs, but only \$38,365 of the \$100,000 accident loss. The lowest value of q in this example at which the insured will still bring a suit is 0.24; at this q, c^* = \$65,101. Thus, even for a suit with the lowest acceptable chance of success, optimal insurance would cover only approximately half of the accident loss in addition to the litigation costs. The prospect of obtaining compensation as a result of a suit makes any higher coverage undesirable.

(iv) The explanation for (c) is the inability to insure against litigation risk, rendering the positive expected return of some suits not worth pursuing. In the example in the preceding paragraph, if $q = 0.8$, then t = \$12,743.[24] In other words, it is desirable to bring suit if and only if \$80,000 – k > \$12,743, which is to say, if and only if k < \$67,257. Hence, for \$67,257 ≤ k < \$80,000, suits would have positive expected value, but would not be optimal to bring.

---

[22] This utility function, which displays decreasing absolute risk aversion, implies, for example, that an individual with income of \$200,000 would be willing to pay up to \$46,154 to avoid a 0.3 probability of a \$100,000 loss, which is to say, an expected loss of \$30,000.

[23] This value (and others presented below) is calculated to two decimal places but is reported only to the nearest dollar.

[24] This number is determined by finding the value of k—call it k^o— at which, given the other parameter values and the optimal choice of c for that k, the individual is indifferent between suing and not suing. Then $t = qh - k^o$.

Proof: (a) If suits are not optimal to bring, then from (2) and $\pi = pc$, the insured's problem is to maximize $(1-p)U(y-pc) + pU(y-pc-h+c)$ over c. Jensen's inequality gives us

$$(1-p)U(y-pc) + pU(y-pc-h+c) \tag{6}$$
$$\leq U((1-p)(y-pc) + p(y-pc-h+c)) = U(y-ph),$$

where the inequality is strict if $y - pc \neq y - pc - h + c$, or if $c \neq h$. And since if $c = h$, expected utility is $U(y-ph)$, (6) implies that $c = h$ is optimal.

(b) Using (3) and $\pi = pc$, we obtain

$$EU_S'(c) = -p(1-p)U'(y-pc) + p(1-p)[qU'(y-pc+c-k) \tag{7}$$
$$+ (1-q)U'(y-pc-h+c-k)],$$

implying that

$$EU_S'(0) = p(1-p)\{[qU'(y-k) + (1-q)U'(y-h-k)] - U'(y)\} > 0, \tag{8}$$

so that $c^* > 0$. Hence, c^* must satisfy

$$EU_S'(c) = -p(1-p)U'(y-pc) + p(1-p)[qU'(y-pc+c-k) \tag{9}$$
$$+ (1-q)U'(y-pc-h+c-k)] = 0,$$

which yields (5). Note also that

$$EU_S''(c) = p^2(1-p)U''(y-pc) + p(1-p)^2[qU''(y-pc+c-k) \tag{10}$$
$$+ (1-q)U''(y-pc-h+c-k)] < 0,$$

which is the second-order condition for a maximum of (3), so that (5) identifies a (unique) maximum. When (5) holds, it must be that

$$y - pc^* - h + c^* - k < y - pc^* < y - pc^* + c^* - k. \tag{11}$$

The first inequality here implies that $c^* < h + k$ and the second that $c^* > k$, so that c^* is in $(k, h+k)$.

(c) We first demonstrate that there is a $\tilde{k}$ satisfying $0 < \tilde{k} < qh$ such that for $k < \tilde{k}$ suit is preferred, at $k = \tilde{k}$ suit and no suit are equally desirable (and by our convention, suit will not occur), and for $k > \tilde{k}$ no suit is preferred. Let $EU_S(k)$ be expected utility when c is chosen optimally given k. At $k = 0$, the insured will be better off if he brings suit than not, for he might win the suit. Hence, $EU_S(0) > U(y - ph)$. At $k = qh$, the insured will be worse off if he brings suit, for

$$(12) \quad EU_S(qh) = (1 - p)U(y - pc) + p[qU(y - pc + c - qh) + (1 - q)U(y - pc - h + c - qh)]$$
$$< (1 - p)U(y - pc) + pU(y - pc + c - h) < U(y - ph).$$

The first inequality follows from the use of Jensen's inequality on the term in brackets, and the second inequality also follows from Jensen's inequality. Since $EU_S(k)$ is continuous and decreasing in k, there must be a unique $\tilde{k}$ satisfying $0 < \tilde{k} < qh$ at which $EU_S(\tilde{k}) = U(y - ph)$ and thus having the claimed properties.

The threshold t in (c) equals $qh - \tilde{k}$. For if $qh - k > qh - \tilde{k}$, or equivalently if $k < \tilde{k}$, suit is optimal to bring; and similarly, if $qh - k \leq qh - \tilde{k}$, or if $k \geq \tilde{k}$, suit is not optimal to bring. Also, because $\tilde{k} = qh - t$ and $0 < \tilde{k} < qh$, we have $0 < qh - t < qh$, or $0 < t < qh$. □

Next consider the optimal insurance policy with a subrogation provision. In this case, if the instruction is $\phi = 0$, not to sue after an accident, the premium constraint will be (1), expected utility will be (2), and by Proposition 1(a), optimal coverage will be $c^* = h$ and utility will be $U(y - ph)$. If, however, $\phi = 1$, to sue after an accident, the premium constraint will be

$$(13) \qquad \pi = pc - p(q(h - s) - k)$$

because the insurer will receive income of $h - s$ if it wins a suit and will incur the litigation cost k if there is an accident. Hence, as is clear from Figure 2, the expected utility of the insured will be

(14) $$EU_S(c, s) = (1 - p)U(y - \pi) + p[(1 - q)U(y - \pi - h + c) + qU(y - \pi - h + c + s)].$$

[Insert Figure 2 here.] The optimal insurance policy with a subrogation provision solves the problem[25]

(15) $$max[U(y - ph), max\ EU_S(c, s) \text{ over } c \text{ and } s].$$

In particular, if the solution is $U(y - ph)$, then the optimal instruction ϕ is not to sue, whereas otherwise it is to sue. We have

Proposition 2. Under the optimal insurance policy with a subrogation provision,

(a) if suits are not optimal to bring, optimal coverage is full, $c^* = h$, and each insured has certain wealth of $y - ph$;

(b) if suits are optimal to bring, optimal coverage is full, $c^* = h$, the optimal subrogation share is zero, $s^* = 0$, and thus each insured has certain wealth of $y - ph + p(qh - k)$; and

(c) suits are optimal to bring if and only if they have positive expected value, $qh - k > 0$, in which case insureds are better off by the positive amount $p(qh - k)$ than if suits are not brought.

Notes. (i) Part (a) is true because the insured faces the standard insurance problem.

(ii) Part (b) holds because it is desirable for the insured to be fully protected against both accident risk and litigation risk, and this is accomplished when coverage is full and the subrogation share is zero.

[25] Note that the structure of the problem in (15) is parallel to that in (4)—in each case, the policy must take into account two courses of action, suit and no suit—but in (4) the decision whether to sue is made ex post by the insured, whereas in (15) this decision is made ex ante by the insured through his instruction ϕ to the insurer.

(iii) Part (c) follows from (b). Because insureds are fully protected against all risks, they will be made better off from suit if and only if suit results in lower premiums. That will be true when suit has positive expected value.

Proof: (a) The proof is identical to that of Proposition 1(a).

(b) By Jensen's inequality, we have

$$EU_S(c, s) = (1 - p)U(y - \pi) + p[(1 - q)U(y - \pi - h + c) + qU(y - \pi - h + c + s)] \tag{16}$$
$$\leq U((1 - p)(y - \pi) + p[(1 - q)(y - \pi - h + c) + q(y - \pi - h + c + s)])$$
$$= U(y - \pi + p(c - h) + pqs).$$

Using (13) we have that $y - \pi + p(c - h) + pqs = y - ph + p(qh - k)$. Hence,

$$EU_S(c, s) \leq U(y - ph + p(qh - k)), \tag{17}$$

where the inequality is strict unless $y - \pi = y - \pi - h + c = y - \pi - h + c + s$, which is to say, unless $c = h$ and $s = 0$. Hence, we know that $c^* = h$ and $s^* = 0$.

(c) If the insurer is instructed not to bring a suit, the insured's wealth is $y - ph$, whereas if the insurer is instructed to bring a suit, the insured's wealth is $y - ph + p(qh - k)$. Consequently, it will be optimal to instruct the insurer to bring a suit if and only if $qh - k > 0$, in which case the insured's wealth will rise by $p(qh - k)$. □

Comparing the previous two propositions, we have[26]

Proposition 3. The optimal insurance policy with a subrogation provision is

(a) equivalent to the optimal pure insurance policy when suits have negative or zero expected value; and

(b) strictly superior to the optimal pure insurance policy when suits have positive expected value.

[26] Shavell (1987, pp. 255-56) shows that subrogation is beneficial because it reduces risk-bearing; but in the model that he studies, suits are costless and the decision to sue is not studied, and s is assumed to equal $h - c$.

Notes. (i) With regard to part (a), we know from Proposition 2(c) that under the optimal insurance policy with a subrogation provision, suits will not be brought when they do not have positive expected value, and hence from Proposition 2(a) that $c^* = h$ and wealth will be certain and equal to $y - pc$. Likewise, from Proposition 1(c), we know that suits will not be brought under the optimal pure insurance policy, and from Proposition 1(a) that the outcome will be the same.

(ii) Part (b) follows because Proposition 2(c) states that under the optimal insurance policy with a subrogation provision, suits will be brought and wealth will be certain and equal to $y - ph + p(qh - k)$ when suits have positive expected value. From Proposition 1, under the optimal pure insurance policy, one of two inferior outcomes will occur when suits have positive expected value: either suits will not be brought, meaning that wealth will be certain but equal only to $y - ph$ because the insurer does not earn the additional expected income $p(qh - k)$ from bringing suit; or suits will be brought, but insureds will bear accident risk (because c^* is generally different from h) as well as litigation risk.

(iii) These two inferior outcomes are illustrated by the example discussed after Proposition 1. There suit will not be brought under the optimal pure insurance policy when k is between \$67,257 and \$80,000, even though such suits would have positive expected value. Thus, for example, if k = \$70,000 and q = 0.8, the insured forgoes additional expected income of \$3,000 by not bringing suit. Though suit would be brought under the optimal pure insurance policy when k is less than \$67,257, the insured will bear accident risk. For instance, if k = \$15,000 and q = 0.8, c^* was seen to be \$53,365, far short of compensating the insured for the accident loss of \$100,000. Moreover, the insured will bear litigation risk: his wealth will vary by \$100,000, depending on whether he wins or loses the suit.

Proof: (a) This follows from Proposition 1(a) and 1(c) and Proposition 2(a) and 2(c).

(b) If $qh - k > 0$, then by Proposition 2(b) and 2(c), a suit will be brought under the optimal insurance policy with a subrogation provision and utility will be $U(y - ph + p(qh - k))$. There are two possible outcomes under a pure insurance policy, depending on whether $qh - k$ exceeds the threshold level t. If $qh - k \le t$, then by Proposition 1(a) and 1(c), a suit will not be brought under a pure insurance policy and utility will be $U(y - ph) < U(y - ph + p(qh - k))$. If $qh - k > t$, then by Proposition 1(b) and 1(c), a suit will be brought under a pure insurance policy and expected utility will be

$$(18)\quad EU_S = (1 - p)U(y - pc^*) + p[qU(y - pc^* + c^* - k) + (1 - q)U(y - pc^* - h + c^* - k)] < U(y - ph + p(qh - k)),$$

where the inequality follows from Jensen's inequality, which must hold strictly because $y - pc^* = y - pc^* + c^* - k = y - pc^* - h + c^* - k$ cannot hold. Hence, again, the insured is worse off than under the optimal insurance policy with a subrogation provision. □

3. Administrative Costs, Moral Hazard, and Non-Monetary Losses

We now consider optimal subrogation provisions given three common factors that alter the nature of optimal insurance coverage: administrative costs, moral hazard, and non–monetary losses.[27] Importantly, each of these factors leads optimal insurance coverage to be less than sufficient to make the insured whole. We demonstrate here that in the cases of administrative costs and moral hazard the optimal subrogation provision provides for the insured to obtain a portion of the proceeds from successful litigation, but not in the case of non-monetary losses.

[27] For simplicity, we do not explain in the present section why an insurance policy with a subrogation provision is superior to a pure insurance policy. The reasons are similar to those discussed in the previous section.

A. Administrative Costs

We suppose here that insurers bear administrative costs that rise with the level of coverage. Notably, the cost to an insurer of verifying the validity of a claim would tend to increase with the magnitude of coverage.[28] We assume for simplicity that such administrative costs rise in proportion to coverage. Let

λ = the administrative cost per dollar of coverage; $\lambda > 0$,

where λ will also be referred to as the *loading factor*.

When the instruction in the policy is $\phi = 0$, not to sue after an accident, the premium constraint is

$$\pi = p(1 + \lambda)c. \tag{19}$$

In this case, expected utility given c is $EU_N(c)$ as stated in (2), where π is determined by (19).

When the instruction is $\phi = 1$, to bring suit if an accident occurs, the premium constraint is

$$\pi = p(1 + \lambda)c - p(q(h - s) - k). \tag{20}$$

Note here that although the loading factor applies to insurance coverage c, we assume that it does not apply to the subrogation payment s to the insured. The justification for this assumption is that, if a suit is won, the insurer would bear essentially no resource cost in making the payment of s—the cost of mailing a check or of wiring funds is negligible. (Of course, litigation costs must be expended to bring a suit, but the resource cost at issue now is that of making a payment to the insured from the judgment, after the suit has been won.) The expected utility of the insured will be $EU_S(c, s)$ as given in (14), with π determined by (20).

---

[28] For example, an insurer would have less to investigate if a homeowner's claim concerned only damage to a stove caused by a small kitchen fire than if the claim concerned a fire that destroyed the entire kitchen—all the appliances, cabinets, flooring, and so forth. On this and other reasons for a loading, see, for example, Harrington and Niehaus (2004, pp. 180-81).

Therefore, the optimal insurance policy with a subrogation provision solves the problem

(21) $$max[max\ EU_N(c) \text{ over } c,\ max\ EU_S(c, s) \text{ over } c \text{ and } s].$$

When we determine the optimal insurance policy using (21), we obtain

Proposition 4. Suppose that there is a positive administrative cost λc associated with the provision of insurance coverage c. Then under the optimal insurance policy with a subrogation provision,

(a) if suits are not optimal to bring, optimal coverage is incomplete, $c^* < h$;

(b) if suits are optimal to bring, optimal coverage is incomplete, $c^* < h$, the optimal subrogation payment is positive, $s^* > 0$, and $c^* + s^* < h$; and

(c) suits are optimal to bring whenever they have positive or zero expected value, $qh - k \geq 0$, and some negative expected value suits are also optimal to pursue.

Notes. (i) When suits are not optimal to bring, c^* is less than full for well-understood reasons: were coverage full, a person would be locally risk-neutral, but because of the loading factor, reducing coverage a small amount would raise his expected wealth.[29]

(ii) When suits are optimal to bring, the reason that c^* is less than full is due not only to the loading factor, on grounds similar to those just explained, but also to the possibility that a suit would be won and that a positive s^* would be obtained, reducing the need for coverage.

(iii) Closely related, the reasons that $s^* > 0$ are essentially these: on one hand, since $c^* < h$, the marginal utility of wealth after an accident and receipt of c^* is higher than in the no-accident state; and on the other hand, the implicit premium that the insured pays to obtain s^* is actuarially fair because, as we observed above, there is no administrative cost associated with the

---

[29] See, for example, Schlesinger (2013, p. 170).

payment of s^* to the insured.[30] In other words, an individual's election of $s^* > 0$ is similar to his purchasing insurance on actuarially fair terms in the event that his suit is won, when his usual insurance coverage was incomplete due to the loading factor.[31]

(iv) However, choosing a positive s^* is not identical to purchasing insurance coverage, for the implicit premium paid for s^* (namely, pqs^*) is borne in each contingency faced by the insured, including when he has an accident and loses his subsequent suit. In that contingency, his wealth is lowest, so the increase in the premium is relatively costly in utility terms. This observation explains why $c^* + s^* < h$ rather than equals h.

(v) That suits are optimal to pursue when they have positive expected value is clear: given any contemplated level of coverage c, premiums can be lowered if such suits are brought, and the insured need not bear any risk since $s = 0$ can be chosen. That suits with zero or small negative expected value are also optimal to bring concerns the point explained in (iii) that, since $c^* < h$, expected utility is raised by bringing suit since that permits $s^* > 0$.

Proof: (a) If $c^* = 0$, then $c^* < h$. If $c^* > 0$, it is determined by the first-order condition from (2), making use of (19),

$$(22)\qquad p(1 + \lambda)[(1 - p)U'(y - \pi) + pU'(y - \pi - h + c)] = pU'(y - \pi - h + c).$$

Since $\lambda > 0$, (22) implies that $(1 - p)U'(y - \pi) + pU'(y - \pi - h + c) < U'(y - \pi - h + c)$, or that $U'(y - \pi) < U'(y - \pi - h + c)$, meaning that $c^* < h$.

(b) We first show that $s^* > 0$. Suppose otherwise, that $s^* = 0$. Then the proof of (a) implies that $c^* < h$; for even though π is determined by (20) rather than (19), (22) still holds.

---

[30] The increase in the premium π due to s^* is pqs^* (see (20)), the expected value of the subrogation payment s^*.

[31] Our result that $s^* > 0$ may be contrasted to the conclusion that we would have reached under the assumption of Sykes (2001, pp. 391-92) that the same loading factor λ is associated with a payment s to an insured as with insurance coverage c. Were that true, s^* would be zero. But the proper assumption is that there is essentially no administrative cost associated with a positive s, as we explained above.

We next demonstrate that $c^* < h$ and $s^* = 0$ leads to a contradiction—which will imply that $s^* > 0$. Observe that

$$EU_S'(s) = -pq(1-p)U'(y-\pi) + pq[(1-pq)U'(y-\pi-h+c+s) - (1-q)pU'(y-\pi-h+c)], \tag{23}$$

so that

$$EU_S'(0) = pq(1-p)[U'(y-\pi-h+c) - U'(y-\pi)] > 0, \tag{24}$$

where the inequality follows from $c^* < h$. Since $EU_S'(0) > 0$, s^* cannot be 0, a contradiction.

Last, we show that $c^* + s^* < h$, which will also imply that $c^* < h$. Now since $s^* > 0$, $EU_S'(s) = 0$ holds, which from (23) is

$$(1-p)U'(y-\pi) + p[qU'(y-\pi-h+c+s) + (1-q)U'(y-\pi-h+c)] = U'(y-\pi-h+c+s). \tag{25}$$

If $c^* + s^* < h$ is not true, then $c + s \geq h$, implying that $y - \pi - h + c + s \geq y - \pi$. Also, since $s > 0$, $y - \pi - h + c + s > y - \pi - h + c$. Thus, the wealth argument on the right-hand side of (25) is greater than or equal to each of the three wealth arguments on the left-hand side, and strictly greater than $y - \pi - h + c$. Hence, $U'(y - \pi - h + c + s)$ is less than or equal to each of the marginal utilities on the left-hand side, and strictly less than $U'(y - \pi - h + c)$. Furthermore, the probability weights on the three marginal utilities on the left-hand side add to 1. Accordingly, the left-hand side of (25) is less than the right-hand side, a contradiction. It follows that $c^* + s^* < h$.

(c) We first show that suits that have positive expected value will be brought. Suppose otherwise, that a suit with $qh - k > 0$ is not brought and coverage c^* is purchased. Expected utility will be

$$EU_N(c^*) = (1-p)U(y-\pi) + pU(y-\pi-h+c^*), \tag{26}$$

where $\pi = p(1 + \lambda)c^*$. But this expected utility can be improved upon with an insurance policy in which suit is brought, coverage is maintained at c^*, and $s = 0$. Then expected utility will be

$$EU_S(c^*, 0) = (1 - p)U(y - \hat{\pi}) + pU(y - \hat{\pi} - h + c^*), \tag{27}$$

where $\hat{\pi} = p(1 + \lambda)c^* - p(qh - k)$. Since $qh - k > 0$, $\hat{\pi} < \pi$ and hence (27) exceeds (26), contradicting the optimality of not bringing suit.

Next consider a suit that has zero expected value. If such a suit is not brought, expected utility will be (26), and again consider a policy in which suit is brought, coverage remains at c^*, and $s = 0$. Then (27) will apply, with $\hat{\pi} = p(1 + \lambda)c^* - p(qh - k) = p(1 + \lambda)c^* = \pi$. Hence, $EU_S(c^*, 0) = EU_N(c^*)$. But by (b), we know that if a suit is brought, $s^* > 0$. Accordingly, $EU_S(c^{**}, s^*) > EU_S(c^*, 0)$, where c^{**} is the optimal c when suits are brought. Thus, $EU_S(c^{**}, s^*) > EU_N(c^*)$, contradicting the optimality of not bringing suit.

That it is optimal to bring some negative expected value suits follows from the preceding paragraph. Because it is strictly optimal to bring suit when $qh - k = 0$, by continuity, it must also be strictly optimal to bring suit if litigation costs are $k + \varepsilon$ for small positive ε. □

B. <u>Moral Hazard</u>

We here consider the problem of moral hazard, in which the incentive of an insured to take care to reduce the risk of an accident is dulled by his receipt of insurance payments. Specifically, we assume that the insured's level of care is not observable by the insurer and that it is non-monetary.[32] Let

x = level of care of an insured; $x \geq 0$; and

$d(x)$ = disutility of care; $d(0) = 0$; $d'(0) = 0$; $d'(x) > 0$ for $x > 0$; $d''(x) > 0$.

[32] This assumption simplifies the analysis because it means that the level of care does not affect the wealth of an individual.

We assume that that there are two mutually exclusive types of accident. The first type is caused by the insured, and its probability can be lowered by his exercise of care. For example, as we said in the introduction, a homeowner can reduce the risk of fires by storing flammables away from heat sources. Let

$p_1(x)$ = probability of a type 1 accident, caused by an insured; $0 < p_1(x) < 1$; $p_1'(x) < 0$; $p_1''(x) > 0$.

The second type of accident is caused by another party, such as when a homeowner hires a contractor whose faulty installation of wiring leads to a fire. We assume that the probability of this type of accident is exogenous.[33] Let

p_2 = probability of a type 2 accident, caused by another party; $p_2 > 0$.

Thus, the total probability of an accident is

$$p(x) = p_1(x) + p_2. \tag{28}$$

If an accident occurs, the insurer is assumed not to have any information about its type, so that the instruction ϕ whether to sue cannot be conditioned on its type. If a suit is brought, however, the court is assumed to be able to determine which type of accident occurred and to award a judgment to the insured if and only if the accident was of the second type, caused by another party.[34]

Suppose that the instruction is $\phi = 0$, not to sue after an accident. Then the insured's expected utility is

---

[33] It will be seen that under this assumption, no moral hazard will be created by making a subrogation payment to the insured when an accident is of type 2. The significance of the assumption is only that the moral hazard associated with type 2 accidents is less than that associated with type 1 accidents; it need not be zero.

[34] We presume that the court does not observe the insured's level of care, and thus that the insurer cannot obtain such information from the court in order to combat moral hazard. The motivation for this assumption is that in reality there are many dimensions of care that a court would not be likely to consider. For instance, suppose a homeowner stored oily rags near his basement furnace and they generated a fire. That outcome would be unlikely to lead to a suit against another party, mooting the possibility that a court would learn that the homeowner took deficient care.

(29) $$EU_N = (1 - p(x))U(y - \pi) + p(x)U(y - \pi - h + c) - d(x),$$

where neither c nor π can depend on x because the insurer is unable to observe it. The optimal insurance policy maximizes EU_N over c and π subject to

(30) $$\pi = p(x)c,$$

and

(31) the insured maximizes EU_N over x, given c and π.

The reason for the latter constraint is that when the insured chooses x, his policy terms c and π have been agreed upon and do not depend on x. The role of moral hazard in this problem is obvious from (31) when $c = h$. Then the insured's expected utility is $EU_N = U(y - \pi) - d(x)$, in which case the insured would choose $x = 0$. If, however, c is less than h, then the insured's exposure to risk will be seen to lead him to choose $x > 0$. Let x_N be the x chosen by the insured under the optimal insurance policy assuming that suit is not brought.

Now assume that $\phi = 1$, that suit will be brought if an accident occurs. Then the insured's expected utility is

(32) $$EU_S = (1 - p(x))U(y - \pi) + p(x)[(1 - q)U(y - \pi - h + c) + qU(y - \pi - h + c + s)] - d(x).$$

The optimal insurance policy maximizes EU_S over c, s, and π subject to the constraints

(33) $$\pi = p(x)c - p(x)(q(h - s) - k),$$

and

(34) the insured maximizes EU_S over x, given c, s, and π.

We observe that q, the probability of winning a suit conditional on an accident having occurred, is $p_2/(p_1(x) + p_2)$, because our assumption is that a suit is won if and only if the accident turns out to have been caused by another party. Because q depends on x we will sometimes write it as $q(x)$. Hence, $p(x)q(x)$, the *unconditional* probability that a judgment will be obtained, is $(p_1(x) + p_2)[p_2/(p_1(x) + p_2)] = p_2$. The explanation is that all accidents caused by other parties will result in liability, whereas no accidents caused by insureds will result in liability. It follows that *the level of care cannot affect the number of judgments* because x influences only the probability of accidents caused by insureds. Relatedly, because $p(x)q = p_2$, $p(x)(1 - q) = p(x) - p_2 = p_1(x)$; that is, the probability of having an accident and then losing at trial is the probability of having an accident that was caused by the insured.

Hence, (32) becomes

$$EU_S = (1 - p(x))U(y - \pi) + p_1(x)U(y - \pi - h + c) + p_2U(y - \pi - h + c + s) - d(x), \tag{35}$$

and (33) becomes

$$\pi = p(x)(c + k) - p_2(h - s). \tag{36}$$

Consequently, the choice of care of insureds affects their premiums only through the coverage amount and litigation costs, not through the return to insurers from bringing suits.

Given (35), the constraint (34) will be shown to be equivalent to the first-order condition

$$-p_1'(x)[U(y - \pi) - U(y - \pi - h + c)] = d'(x); \tag{37}$$

that is, the marginal benefit from reducing the risk of type 1 accidents (and thus bearing $h - c$) equals the marginal cost of care. Note that the subrogation payment s does not appear in (37) and thus does not directly affect the insured's choice of care; this is because s is obtained only if there is a judgment, and the probability p_2 of a judgment is not affected by his care.

We can now describe the optimal insurance policy.

Proposition 5. Suppose that moral hazard is associated with the provision of insurance coverage. Then under the optimal insurance policy with a subrogation provision,

(a) if suits are not optimal to bring, optimal coverage is incomplete, $c^* < h$;

(b) if suits are optimal to bring, optimal coverage is incomplete, $c^* < h$, and the optimal subrogation payment is positive, $s^* > 0$; and

(c) suits are optimal to bring whenever they would have positive or zero expected value if suits were not brought, $q(x_N)h - k \geq 0$, and some negative expected value suits are also optimal to pursue.

Notes. (i) When suits are not optimal to bring, the insurance policy becomes an example of the standard insurance policy involving moral hazard. Hence, $c^* < h$ is optimal.[35] The intuition is that if coverage were full and c was lowered marginally, creating a small exposure to risk of $h - c$, there would be a first-order increase in the level of care x and thus a first-order decrease in the premium per dollar of coverage. Moreover, this small exposure to risk would have no first-order negative effect on expected utility due to risk-bearing (because insureds begin from a position of no-risk bearing). Accordingly, some reduction of c from h must be optimal.

(ii) When suits are optimal to bring, the result that $c^* < h$ is essentially due to the argument just given. The explanation for $s^* > 0$ is as follows. First, since $c^* < h$, the marginal utility of wealth after an accident and receipt of c^* is higher than in the no-accident state, meaning that arranging for a positive s is similar to arranging for a beneficial increase in insurance coverage. Second, the possible countervailing argument that receipt of a payment s from the insurance company would dilute the incentives of the insured to take care is not applicable: as we observed (after (37)), the insured's receipt of s is independent of his choice of care x because s is obtained only when an accident is caused by another party.[36] Consequently,

[35] See, for example, the valuable synthesis by Winter (2013, pp. 208-12).

raising s from zero must increase expected utility by reducing the risk-bearing created by $c^* < h$ and without contributing to moral hazard.[37]

(iii) That suits are optimal to pursue when they have positive expected value is based on logic similar to that provided after Proposition 4: given any contemplated level of coverage c, premiums can be lowered if such suits are brought, and the insured need not bear any additional risk since $s = 0$ can be chosen. This suggests that insureds must be better off, but the effect of a lower premium on the optimally chosen level of care needs to be taken into account to make this argument. That suits with zero or small negative expected value are optimal to bring is explained by the point made in note (ii) that, since $c^* < h$, expected utility is raised by bringing suit because that permits $s^* > 0$.

Proof: See the Appendix.

C. Non-Monetary Losses

In this section we consider optimal insurance and subrogation when the losses from an accident include both a monetary component and a non-monetary component, such as when the victim of an accident both bears medical expenses and suffers pain. Let

u = loss of utility due to the non-monetary component of harm; $u > 0$; and

a = court award for the non-monetary loss; $a > 0$.

---

[36] Actually, arranging for $s > 0$ turns out to *increase* care. The reason is that when $s > 0$, the premium rises, and an increase in the premium will elevate care (because the premium increase is equivalent to a reduction in the insured's wealth). See the demonstration that $x'(s) > 0$ in step (v) of the proof of part (b).

[37] Our result that $s^* > 0$ would be different under the assumption of Sykes (2001, pp. 393-94), who supposes that the problem of moral hazard is identical whether an accident results in a successful suit or does not. If that were true, s^* would equal zero. But it is implausible that the fact that a suit is successful would provide no information to an insurer about the role of the insured in an accident (in our case, about whether the accident was caused by the insured or by another party).

Thus, if a person is involved in an accident and his ultimate level of wealth is z, his utility will be $U(z) - u$.[38] In all other respects, the model analyzed in this subsection is that employed in Section 2, including that the monetary loss is h and that the court will award h for this component of loss.

If $\phi = 0$, so that no suit will be brought after an accident, the premium will be (1) and the insured's expected utility given c will be

(38) $$EU_N(c) = (1 - p)U(y - \pi) + p[U(y - \pi - h + c) - u].$$

If $\phi = 1$, so that a suit will be brought after an accident, the premium will be

(39) $$\pi = pc - p(q(h + a - s) - k)$$

because the insurer's expected profit given that an accident occurs is $q(h + a - s) - k$. Hence, the insured's expected utility given c and s will be

(40) $$EU_S(c, s) = (1 - p)U(y - \pi) + p[(1 - q)U(y - \pi - h + c) + qU(y - \pi - h + c + s) - u].$$

The optimal insurance policy with a subrogation provision solves

(41) $$max[max\ EU_N(c) \text{ over } c, max\ EU_S(c, s) \text{ over } c \text{ and } s].$$

We have

Proposition 6. Suppose that there are both monetary and non-monetary losses when an accident occurs. Then under the optimal insurance policy with a subrogation provision,

(a) if suits are not optimal to bring, optimal coverage equals the monetary loss alone, $c^* = h$;

(b) if suits are optimal to bring, optimal coverage equals the monetary loss alone, $c^* = h$, and the optimal subrogation share is zero, $s^* = 0$; and

---

[38] Accordingly, the occurrence of the non-monetary loss does not affect his marginal utility of wealth. This assumption seems most realistic, though cases in which non-monetary losses affect the marginal utility of wealth are considered in the literature on insurance; see originally Arrow (1974) and Cook and Graham (1977).

(c) suits are optimal to bring if and only if they have positive expected value, $q(h + a) - k > 0$.

Notes: (i) The result in part (a) is the well-understood point that optimal insurance coverage for a non-monetary loss is zero when such a loss does not affect the marginal utility of wealth.[39]

(ii) The result in part (b) is due to two factors: on one hand, optimal insurance coverage is limited to monetary losses h as just observed; on the other hand, the insured can forgo collecting a positive subrogation share but still benefit from the court award a for the non-monetary loss through a lower premium.[40]

(iii) The result in part (c) is due to the point that the insurer's income from suits, and thus the insured's reduction in his premium, will be maximized if and only if all positive expected value suits are pursued.

Proof: (a) We want to show that

$$EU_N(h) = U(y - ph) - pu > (1 - p)U(y - pc) + pU(y - pc - h + c) - pu = EU_N(c) \tag{42}$$

for $c \neq h$. This is equivalent to $U(y - ph) > (1 - p)U(y - pc) + pU(y - pc - h + c)$ for $c \neq h$, which follows from Jensen's inequality.

(b) We want to show that

$$\begin{aligned} EU_S(h, 0) = U(y - ph + p(q(h + a) - k)) - pu > (1 - p)U(y - \pi) + p[(1 - q)U(y - \pi - h + c) \\ + qU(y - \pi - h + c + s) - u] = EU_S(c, s) \end{aligned} \tag{43}$$

for any $c \neq h$ or any $s > 0$. Equivalently, we want to show that

$$U(y - ph + p(q(h + a) - k)) > \tag{44}$$

[39] See Arrow (1974) and Cook and Graham (1977).

[40] In contrast, Sykes (2001, p. 389) finds that it is optimal for the insured to retain the entire award *a* for non-monetary losses. But that is because Sykes assumes that it is impossible for the insured—who brings suit himself—to transfer the award *a* to the insurer. There is no apparent economic justification for this assumption.

$$(1-p)U(y-\pi) + p[(1-q)U(y-\pi-h+c) + qU(y-\pi-h+c+s)]$$

for any $c \neq h$ or any $s > 0$. Now expected wealth on the right-hand side of this inequality is

$$(1-p)(y-\pi) + p[(1-q)(y-\pi-h+c) + q(y-\pi-h+c+s)] = \tag{45}$$

$$y-\pi-p(h-c)+pqs = y-[pc-p(q(h+a-s)-k)]-p(h-c)+pqs$$

$$= y-ph+p(q(h+a)-k,$$

which is the wealth on the left-hand side. Hence Jensen's inequality demonstrates that (44) holds, and strictly because the levels of wealth $y-\pi$, $y-\pi-h+c$, and $y-\pi-h+c+s$ are not all equal if $c \neq h$ or if $s > 0$.

(c) We know from the preceding steps that $EU_N(c^*) = U(y-ph) - pu$ and that $EU_S(c^*, s^*) = U(y-ph+p(q(h+a)-k)) - pu$. Thus, suit will be desirable if and only if

$$U(y-ph+p(q(h+a)-k)) > U(y-ph), \tag{46}$$

which is to say if and only if $p(q(h+a)-k) > 0$ or, equivalently, $q(h+a)-k > 0$. □

4. <u>Discussion</u>

We conclude with observations about several issues that we did not consider above.

Combating moral hazard as a reason for subrogation. In the absence of subrogation provisions, an accident victim can profit from an accident, for he can collect both insurance coverage from his insurer and damages from a liable party. Hence, the insured might have little incentive to avoid accidents and even wish to foment them. This potentially severe problem of moral hazard is alleviated by an insurance policy with a subrogation provision, for under it the insured would not receive much or any of the proceeds from suit and thus not actually benefit from the occurrence of accidents.[41] Accordingly, as some commentators have observed,[42] subrogation can be explained in part as a mechanism for combating a serious moral hazard.[43]

Cooperation of insureds in the litigation process. When suit is desirable under the terms of subrogation provisions, the cooperation of insureds—notably, supplying information needed in the litigation process and giving testimony—will be important to achieving a successful outcome. This suggests not only that optimal subrogation provisions would include clauses requiring such cooperation—which they do in fact[44]—but also that it would be beneficial for the subrogation share paid to insureds to be positive to better motivate them to cooperate.

Asymmetric information and delegation of the decision to litigate. Although we assumed that both insureds and insurers know the probability of winning at trial q and litigation costs k, in reality insureds especially will tend to have imperfect information about these variables. Thus, the instructions in the insurance policy concerning litigation cannot depend on q and k. If as a result the insured delegates the litigation decision to the insurer, the insurer may or may not bring suit when the insured would wish. In the basic case in Section 2 and in the case of non-monetary losses, the insured's and the insurer's interests will be aligned. In those two cases the insured desires that the subrogation payment s is zero and that suits are brought if and only if they have positive expected value. Clearly, if s is zero—meaning that the insurer keeps the entire proceeds from litigation—the insurer will decide to bring suits exactly when they have positive expected

[41] This issue could not have arisen in our particular model of moral hazard because we assumed that when an insured played a role in an accident, he could not win a suit.

[42] See, for example, Harrington and Niehaus (2004, p. 195), Kimball and Davis (1962, p. 869), and Sykes (2001, p. 384).

[43] That is, the desire of insureds to have subrogation provisions can be explained in part because subrogation lowers accident risks and thus premium rates by ameliorating a strong moral hazard.

[44] See Abraham and Schwarcz (2015, p. 208) for a typical example.

value. However, in the cases of loading and moral hazard, there will be a conflict of interest between the insured and the insurer. For then the insured desires that s is positive and that suits are brought if they have positive expected value or a small negative expected value. But if s is positive, the insurer will not capture the entire judgment and therefore will not bring some positive expected value suits and of course will not bring any negative expected value suits. In the light of such conflicts of interest between insureds and insurers, it is not surprising that subrogated insurers face general duties not to unduly compromise the well-being of insureds in litigation.[45]

Legal restrictions on subrogation. Subrogation provisions may be constrained by law,[46] and when so the welfare of insureds may be adversely affected.[47] For example, the law bars subrogated insurers from collecting awards for pain and suffering.[48] This limitation reduces the expected utility of insureds because, as we discussed in Section 3.C, insureds would prefer not to receive any compensation for pure losses in utility; they would prefer instead to obtain lower premiums that subrogation of pain and suffering awards would generate. Another example is that the law precludes subrogation of awards for wrongful death[49] even though insureds would presumably wish their life insurers to have subrogation rights for the general reasons discussed in Section 2.[50]

---

[45] See Jerry and Richmond (2012, pp. 156-57).

[46]See Dobbyn and French (2016, pp. 432-34) and Jerry and Richmond (2012, pp. 651-52).

[47] On this general theme see Reinker and Rosenberg (2007).

[48] In particular, courts generally permit subrogated insurers to collect damages only for types of losses that were insured; see, for example, Couch on Insurance (2015, §223: 85). Thus, because insurance coverage against pain and suffering is essentially non-existent, subrogated insurers would be barred from retaining pain and suffering damages.

[49] See, for example, Jerry and Richmond (2012, pp. 653-54).

The "made whole" rule of subrogation. According to an important default rule of the law on subrogation, an insured who has not been fully compensated for his loss by his insurance coverage should be given a share of litigation proceeds sufficient to make him whole.[51] Only after the insured is made whole is the insurer permitted to retain any litigation proceeds. Our results that under loading and moral hazard it is optimal for insureds to receive a positive portion of damages are roughly consistent with the foregoing made whole rule, for in those cases optimal insurance coverage is less than full. However, our result that when losses are non-monetary the entire judgment should be retained by the insurer implies that the made whole rule would be undesirable to apply in that case.

Selling claims as an alternative to subrogation. It might be thought that many of the virtues of subrogation could be obtained by insureds by purchasing pure insurance policies and then selling their claims after accidents occurred to other parties or directly to their insurers. In that way, insureds would obtain value for their claims without bearing litigation risk. However, sales of claims are not observed in fact (and we did not consider such sales in our analysis). The explanation appears to lie in several factors. First, legal barriers to the sale of claims exist.[52] Second, costs of obtaining information about claims would be borne by potential purchasers of claims, for to assess the value of claims, they would have to learn about the magnitude of the loss, the likelihood of winning, and the roles of the insured and of the injurer in the accident in question. In contrast, the insurer would already have much of this information in order to fulfill its insurance functions, including the control of moral hazard. Hence, the insurer would enjoy a

---

[50] Life insurance may be viewed as insurance for an insured's beneficiary (say a spouse) against the cessation of the insured's stream of income. As such, the lesson from the analysis in Section 2 is that subrogation rights for life insurers would, if permitted, benefit insureds by avoiding overinsurance against loss of the insured's stream of income—resulting in a lower premium—and reducing or eliminating the bearing of litigation risk.

[51] On this "made whole" rule, see, for example, Jerry and Richmond (2012, pp. 654-57).

[52] See, for example, American Jurisprudence (2016, Assignments §§53, 55) and Sebok (2011, pp. 74-121).

natural cost advantage over unrelated buyers of claims. Third, risk-bearing would not be eliminated through the sale of claims because the prices received by insureds would be variable, depending on whether an insured had a suit that he could bring and, if so, on its expected value.[53]

Subrogation versus insurance contingent on the outcome of litigation. One might believe that the benefits of subrogation could be achieved under an insurance policy that specifies when the insured will bring suit and, if a suit is brought, that makes an insurance payment only after it is resolved. Specifically, suppose that such a policy dictates that the insured will bring a suit if and only if it has positive expected value, and that insurance coverage equals litigation costs plus harm if suit is brought and lost, just litigation costs if suit is brought and won, and only harm if suit is not brought. Under this policy, the final levels of wealth of the insured would be the same as those under subrogation. However, such an arrangement would not in fact be desirable for insureds because, as we explained in Section 2, they wish to receive compensation immediately after an accident (because of their consumption needs at that time).

Subrogation and social welfare. Although our focus has been on subrogation as a feature of insurance contracts that maximize the expected utility of insureds, subrogation also has broader effects that influence social welfare. Notably, subrogation results in a greater volume of suit against potentially liable parties (compare Proposition 1(c) to Proposition 2(c)), and thus increases the deterrence of undesirable acts as well as the litigation costs that society incurs. An analysis of the effects of subrogation on social well-being would take such factors into account.[54]

---

[53] In theory, this problem could be cured by having the insured sell his right to bring suits before any accident has occurred. That kind of sale, however, would still suffer from an informational disadvantage relative to subrogation.

[54] Gomez and Penalva (2015) undertake such an analysis in comparing a subrogation regime, a regime barring subrogation, and a regime in which insurance benefits are subtracted from damage payments. See also the remarks in Abraham (1986, pp. 154-55).

Appendix

Proof of Proposition 5: (a) We demonstrate that $c^* < h$ when $\phi = 0$ in several steps.

(i) c^* cannot exceed h: From (29), we have

(A1) $$EU_N'(x) = -p_1'(x)[U(y-\pi) - U(y-\pi-h+c)] - d'(x),$$

since $p'(x) = p_1'(x)$. If $c \geq h$, (A1) is negative for $x > 0$; thus, $x = 0$ would be chosen by the insured. Hence, using (30), we know that for any $c > h$,

(A2) $$EU_N = (1-p(0))U(y-p(0)c) + p(0)U(y-p(0)c-h+c) - d(0),$$

and for $c = h$,

(A3) $$EU_N = U(y-p(0)h) - d(0).$$

Jensen's inequality implies that (A3) exceeds (A2), so $c > h$ cannot be optimal.

(ii) For any c in $[0, h]$ and any π, the insured's choice of x is uniquely determined by the first-order condition $EU_N'(x) = 0$: If $c < h$, (A1) implies that $EU_N'(0) > 0$, so that $x > 0$ must be optimal. Hence, $EU_N'(x) = 0$ must hold, that is

(A4) $$-p_1'(x)[U(y-\pi) - U(y-\pi-h+c)] = d'(x).$$

If $c = h$, we noted in step (i) that $x = 0$ is optimal; and when this is the case, (A4) also holds. Thus, we know that for any c in $[0, h]$ and π, the chosen x must satisfy (A4). We also note that if x satisfies (36) the second-order condition for a local maximum is satisfied.[55] To see that the solution to (A4) is unique, rewrite (A4) as

(A5) $$[U(y-\pi) - U(y-\pi-h+c)] = -d'(x)/p_1'(x).$$

Since $-d'(x)/p_1'(x)$ is strictly increasing in x, the solution to (A4) must be unique.[56]

[55] $EU_N''(x) = -p_1''(x)[U(y-\pi) - U(y-\pi-h+c)] - d''(x) < 0$ for any $c \leq h$.

[56] In other words, we have justified the use of the first-order approach in describing the choice x of the insured given his policy. On this issue, see, for example, Rogerson (1985).

(iii) For any c in $[0, h]$, the insured's choice of x is uniquely determined by the first-order condition (A4) and by (30): In other words, the claim is that this x is uniquely determined by the condition[57]

$$\text{(A6)} \qquad -p_1'(x)[U(y - p(x)c) - U(y - p(x)c - h + c)] = d'(x).$$

We know from step (ii) and (30) that (A6) must hold at any x that is chosen given c. To show that the x solving (A6) is unique, note first that if $c = h$, then $x = 0$ is obviously the unique solution to (A6). Now assume that $c < h$ and suppose to the contrary that (A6) holds for some $x_1 < x_2$. Then

$$\text{(A7)} \qquad U(y - p(x_1)c) - U(y - p(x_1)c - h + c) > U(y - p(x_2)c) - U(y - p(x_2)c - h + c),$$

since $p(x_1) > p(x_2)$ and U is concave. We can rewrite (A6) as

$$\text{(A8)} \qquad U(y - p(x)c) - U(y - p(x)c - h + c) = -d'(x)/p_1'(x).$$

Using (A8), (A7) implies that

$$\text{(A9)} \qquad -d'(x_1)/p_1'(x_1) > -d'(x_2)/p_1'(x_2).$$

Yet (A9) cannot hold because, as we observed above, $-d'(x)/p_1'(x)$ is strictly increasing in x, a contradiction. Let us denote the x that solves (A6) given c by $x(c)$.

(iv) $c^* < h$: We have shown that the problem of maximizing (29) over c subject to (30) and (31) is equivalent to maximizing

$$\text{(A10)} \qquad EU_N(c) = (1 - p(x(c)))U(y - p(x(c))c) + p(x(c))U(y - p(x(c))c - h + c) - d(x(c))$$

over $c \in [0, h]$. Using (A6), we have

$$\text{(A11)} \qquad EU_N'(c) = -[p'(x(c))x'(c)c + p(x(c))][(1 - p(x(c)))U'(y - p(x(c))c) + p(x(c))U'(y - p(x(c))c - h + c)] + p(x(c))U'(y - p(x(c))c - h + c).$$

[57] Note that the uniqueness of the solution to (A6) does not follow from (ii). Suppose that for some c, there exist $x_1 < x_2$ that each satisfy (A6). That would not be inconsistent with (ii), since π would be $p(x_1)c$ for x_1 and $p(x_2)c$, a different value, for x_2.

Hence,

(A12) $$EU_N'(h) = -p'(0)x'(h)hU'(y - p(0)h).$$

To prove that $c^* < h$, we show that $EU_N'(h) < 0$, which will hold if $x'(h) < 0$. To demonstrate the latter, we implicitly differentiate (A6) with respect to c and solve for $x'(c)$. The sign of $x'(c)$ must equal the sign of[58]

(A13) $$\partial\{-p'(x)[U(y - p(x)c) - U(y - p(x)c - h + c)] - d'(x)\}/\partial c,$$

which is

(A14) $$p'(x)[p(x)U'(y - p(x)c) + (1 - p(x))U'(y - p(x)c - h + c)].$$

At $c = h$, (46) is $p'(x)U'(y - p(x)h) < 0$. Thus, $c^* < h$.

(b) We demonstrate that $c^* < h$ and $s^* > 0$ when $\phi = 1$ in several steps.

(i) c^* cannot exceed h: Using (35), we have

(A15) $$EU_S'(x) = -p_1'(x))[U(y - \pi) - U(y - \pi - h + c)] - d'(x).$$

If $c \geq h$, the term in brackets is less than or equal to zero, so that $EU_S'(x) < 0$ for $x > 0$. Hence, the insured would choose $x = 0$ for $c \geq h$. Thus, for any $c > h$,

(A16) $$EU_S(c, s) = (1 - p(0))U(y - \pi(c, s)) + p_1(0)U(y - \pi(c, s) - h + c)$$
$$+ p_2U(y - \pi(c, s) - h + c + s);$$

and for $c = h$ and $s = 0$,

(A17) $$EU_S(h, 0) = (1 - p(0))U(y - \pi(h, 0)) + p_1(0)U(y - \pi(h, 0)) + p_2U(y - \pi(h, 0))$$
$$= U(y - p(0)h).$$

But it can readily be verified that $U(y - p(0)h)) > EU_S(c, s)$ by Jensen's inequality. This means that a policy with $c^* > h$ could not be optimal.

---

[58] Equation (A6) is of the form $F(x, c) = 0$, where F is the first-order condition for the optimal choice of x by the insured. Implicitly differentiating, we obtain $F_x x'(c) + F_c = 0$, so that $x'(c) = -F_c/F_x$. We know that $F_x < 0$, for this is the second-order condition for x to have been chosen optimally. Hence, the sign of $x'(c)$ equals the sign of F_c.

(ii) For any c in $[0, h]$ and any s and π, the insured's choice of x is uniquely determined by the first-order condition (37), $EU_S'(x) = 0$: If $c < h$, (A15) implies that $EU_S'(0) > 0$, so that $x > 0$ is optimal and $EU_S'(x) = 0$ must hold. If $c = h$, we observed in step (i) that $x = 0$, and it is clear that $EU_S'(0) = 0$ holds. We also note that if x satisfies (37), the second-order condition for a local maximum is satisfied. That the solution is unique follows from the argument made in step (ii) of the proof of part (a).

(iii) For any c in $[0, h]$, the insured's choice of x is uniquely determined by the first-order condition (37) and by (36): That is, x is uniquely determined by

$$\text{(A18)} \quad -p_1'(x)[U(y - (p(x)(c + k) - p_2(h - s))) - U(y - (p(x)(c + k) - p_2(h - s)) - h + c)] = d'(x).$$

This claim follows from an argument essentially identical to that made in step (iii) of the proof of part (a). We continue to let $x(c)$ denote the insurer's choice of x given c (suppressing s in the notation).

(iv) $c^* < h$: Having shown in step (i) that $c^* \leq h$, we can prove that $c^* < h$ if we demonstrate that $c^* = h$ is not possible. To this end, we first observe that if $c^* = h$, it must be that $s^* = 0$. In particular, if $c = h$, we know from step (i) that $x = 0$. Hence, if $c = h$ and $s > 0$, we have

$$\text{(A19)} \qquad EU_S(h, s) = (1 - p(0))U(y - \pi(h, s)) + p_1(0)U(y - \pi(h, s)) + p_2U(y - \pi(h, s) + s),$$

whereas from (A17) we have $EU_S(h, 0) = U(y - p(0)h))$. Since $U(y - p(0)h)) > EU_S(h, s)$ by Jensen's inequality, $c = h$ and $s > 0$ cannot be optimal.

Finally, we show that the conclusion that $c^* = h$ and $s^* = 0$ leads to a contradiction—that the insured would be better off if $c < h$ when $s = 0$. We have

$$\text{(A20)} \qquad EU_S(c, 0) = (1 - p(x(c)))U(y - \pi(c, 0)) + p(x(c))U(y - \pi(c, 0) - h + c) - d(x(c)),$$

so that, using (37),

(A21) $$dEU_S(c, 0)/dc = -(1 - p(x(c))(p'(x(c))x'(c)(c + k) + p(x(c)))U'(y - \pi(c, 0))$$

$$+ (1 - p(x(c)) - p'(x(c))x'(c)(c + k))p(x(c))U'(y - \pi(c, 0) - h + c).$$

Hence,

(A22) $$dEU_S(h, 0)/dc = -p'(x(h))x'(h)(h + k)U'(y - \pi(h, 0)).$$

Because $x'(h) < 0$,[59] $dEU_S(h, 0)/dc < 0$, meaning that c^* must be less than h.

(v) $s^* > 0$: We now hold c constant at $c^* < h$ and show that if $s = 0$, expected utility can be raised by increasing s, implying that $s^* > 0$. To this end, let us treat x as a function of s, determined implicitly by (37) and (36), and π as a function of s determined by (36). Hence, we may write EU_S as $EU_S(x(s), \pi(s), s)$. Thus, our object is to demonstrate that $dEU_S(x(0), \pi(0), 0)/ds > 0$. Using (35) and (37), we have

(A23) $$dEU_S(x(s), \pi(s), s)/ds = -\pi'(s)\{(1 - p(x(s)))U'(y - \pi(s)) +$$

$$p_1(x(s))U'(y - \pi(s) - h + c^*) + p_2U'(y - \pi(s) - h + c^* + s)\} + p_2U'(y - \pi(s) - h + c^* + s).$$

Observe that $\pi'(s) = p'(x(s))x'(s)(c^* + k) + p_2$. Hence, at $s = 0$, (A23) becomes

(A24) $$-p'(x(0))x'(0)(c^* + k)[(1 - p(x(0)))U'(y - \pi(0)) + p(x(0))U'(y - \pi(0) - h + c^*)]$$

$$+ p_2(1 - p(x(0)))[U'(y - \pi(0) - h + c^*) - U'(y - \pi(0))].$$

Since $c^* < h$, the second term in (A24) is positive. Thus, a sufficient condition for (A24) to be positive is that the first term is positive, which will be true if $x'(0) > 0$. To see that this holds, substitute (36) into (37) and implicitly differentiate the resulting expression with respect to s and solve for $x'(s)$. The sign of $x'(s)$ must equal the sign of[60]

[59] The sign of $x'(c)$ equals the sign of the partial derivative with respect to c of $-p_1'(x)[U(y - \pi) - U(y - \pi - h + c)] - d'(x)$, using the logic explained in note 58. Since $\pi = p(x)(c + k) - p_2h$, the partial derivative with respect to c is $-p_1'(x)p(x)[U'(y - \pi - h + c) - U'(y - \pi)] + p_1'(x)U'(y - \pi - h + c)$. At $c = h$, this becomes $p_1'(x)U'(y - \pi) < 0$.

[60] See note 58 above.

(A25) $$\partial\{-p'(x)[U(y - p(x)(c^* + k) + p_2(h - s))$$

$$- U(y - p(x)(c^* + k) + p_2(h - s) - h + c^*)] - d'(x)\}/\partial s,$$

which is

(A26) $$p'(x)p_2[U'(y - p(x)(c + k) + p_2(h - s)) - U'(y - p(x)(c + k) + p_2(h - s) - h + c)] > 0,$$

where the inequality follows from observing that the bracketed term is negative because $c^* < h$. Thus, $x'(0) > 0$ and $s^* > 0$.

(c) Again, we prove the claim in steps.

(i) If $q(x_N)h - k > 0$, suits are optimal to bring: Suppose to the contrary that suits are not optimal to bring and let c_N^* be the optimal c in this case. We will show that there is an insurance policy under which suits are brought and insureds are better off, which will be a contradiction. In particular, hold c fixed at c_N^* and let

(A27) $$s = (q(x_N)h - k)/q(x_N).$$

Then $s > 0$ by our hypothesis and, using (33),

(A28) $$\pi(c_N^*, s) = p(x_N)c_N^* - p(x_N)(q(x_N)(h - s) - k) = p(x_N)c_N^*.$$

Because c and π have not changed, it is clear from (37) and (A4) that x_N will continue to be chosen by the insured. The insured must be better off under this policy because he has the same coverage, pays the same premium, chooses the same level of care, but obtains $s > 0$ whenever he wins a suit, which occurs with probability $p_2 > 0$. Hence, it must be optimal for suit to be brought.

(ii) If $q(x_N)h - k = 0$, suits are optimal to bring: Suppose to the contrary that suits are not optimal to bring. We will show that there is an insurance policy under which suits are brought and insureds are better off. Again, hold c fixed at c_N^* and let $s = 0$. Then $\pi(c_N^*, s) = \pi(c_N^*, 0) = p(x_N)c_N^*$. Because c and π have not changed, x_N will still be chosen. Hence, the insured is just as well off as he was when suits were not brought. However, by part (b) of the proposition, $s^* > 0$. Thus, the policy with c_N^* and $s = 0$ cannot be optimal, meaning that there exists a policy involving suit that is superior, and thus better than the optimal policy involving no suit.

(iii) Some suits for which $q(x_N)h - k < 0$ are optimal to bring: Let k_o be such that $q(x_N)h - k_o = 0$. We claim that for all k in $(k_o, k_o + \varepsilon)$, for a sufficiently small positive ε, suits are optimal to bring. This will prove the claim since for such k, $q(x_N)h - k < 0$. Let $EU_S(c^*, s^*, k)$ denote expected utility when suit is brought and c and s are chosen optimally given k. Also, observe that $EU_N(c_N^*)$ does not depend on k. Now in (ii) we showed that $EU_S(c^*, s^*, k_o) > EU_N(c_N^*)$. And since $EU_S(c^*, s^*, k)$ is continuous in k, we must have $EU_S(c^*, s^*, k) > EU_N(c_N^*)$ for all k within ε of k_o for a sufficiently small positive ε. Hence, suits are optimal to bring for k in $(k_o, k_o + \varepsilon)$. □

References

Abraham, Kenneth S. 1986. *Distributing Risk*: *Insurance, Legal Theory, and Public Policy*. New Haven: Yale.

Abraham, Kenneth S., and Daniel Schwarcz. 2015. *Insurance Law and Regulation*. *Cases and Materials*. 6th ed. New York: Foundation.

American Jurisprudence, 2nd ed. 2016. St. Paul: Thomson Reuters.

Arrow, Kenneth J. 1974. "Optimal Insurance and Generalized Deductibles." *Scandinavian Actuarial Journal*. 1-42.

Baker, Tom. 2008. *Insurance Law and Policy*. 2nd ed. Frederick: Aspen.

Bleed, Lisa N. 2001. "Enforcing Subrogation Provisions as 'Appropriate Equitable Relief' Under ERISA Section 502(A)(3)." *University of San Francisco Law Review*, 35: 727-750.

Cook, Philip J., and Daniel A. Graham. 1977. "The Demand for Insurance and Protection: The Case of Irreplaceable Commodities." *Quarterly Journal of Economics*, 91: 143-156.

Couch on Insurance. 2015.

Dionne, Georges, Ed. 2013. *Handbook of Insurance.* 2nd ed. New York: Springer.

Dobbyn, John F., and Christopher C. French. 2016. *Insurance Law*. 5th ed. St. Paul: West.

Gomez, Fernando, and Jose Penalva. 2015. "Tort reform and the theory of coordinating tort and insurance." *International Review of Law and Economics*, 43: 83-97.

Harrington, Scott E., and Gregory R. Niehaus. 2004. *Risk Management and Insurance*. 2nd ed. New York: McGraw-Hill.

"International Comparative Review of Subrogation Law." *Insurance Day*, September 2009.

Jerry II, Robert H., and Douglas R. Richmond. 2012. *Understanding Insurance Law*. 5th ed. New Providence: Matthew Bender.

Kimball, Spencer L., and Don A. Davis. 1962. “The Extension of Insurance Subrogation.” *Michigan Law Review*, 60: 841-872.

Marasinghe, M. L. 1976a. “An Historical Introduction to the Doctrine of Subrogation: The Early History of the Doctrine I.” *Valparaiso University Law Review*, 10:45-65.

Marasinghe, M. L. 1976b. “An Historical Introduction to the Doctrine of Subrogation: The Early History of the Doctrine II.” *Valparaiso University Law Review*, 10:276-299.

“Money machine: subrogation strategy is evolving in the US.” *Insurance Day*, May 2015.

Reinker, Kenneth S., and David Rosenberg. 2007. “Unlimited Subrogation: Improving Medical Malpractice Liability.” *Journal of Legal Studies*, 36: S261-289.

Rogerson, William P. 1985. “The First-Order Approach to Principal-Agent Problems.” *Econometrica*, 53: 1357-1367.

Schlesinger, Harris. 2013. “The Theory of Insurance Demand.” Ch. 7 in Georges Dionne (ed.), *Handbook of Insurance*, 2nd ed., New York: Springer.

Sebok, Anthony J. 2011. “The Inauthentic Claim.” *Vanderbilt Law Review*, 64: 61-139.

Shavell, Steven. 1987. *Economic Analysis of Accident Law*. Cambridge: Harvard.

Sykes, Alan. 2001. “Subrogation and Insolvency.” *Journal of Legal Studies*, 30: 383-399.

Viscusi, W. Kip 1989. “The Interaction between Product Liability and Workers’ Compensation as Ex Post Remedies for Workplace Injuries.” *Journal of Law, Economics, & Organization*, 5: 185-210.

Winter, Ralph A. 2013. “Optimal Insurance Contracts Under Moral Hazard.” Ch. 9 in Georges Dionne (ed.), *Handbook of Insurance*, 2nd ed., New York: Springer.

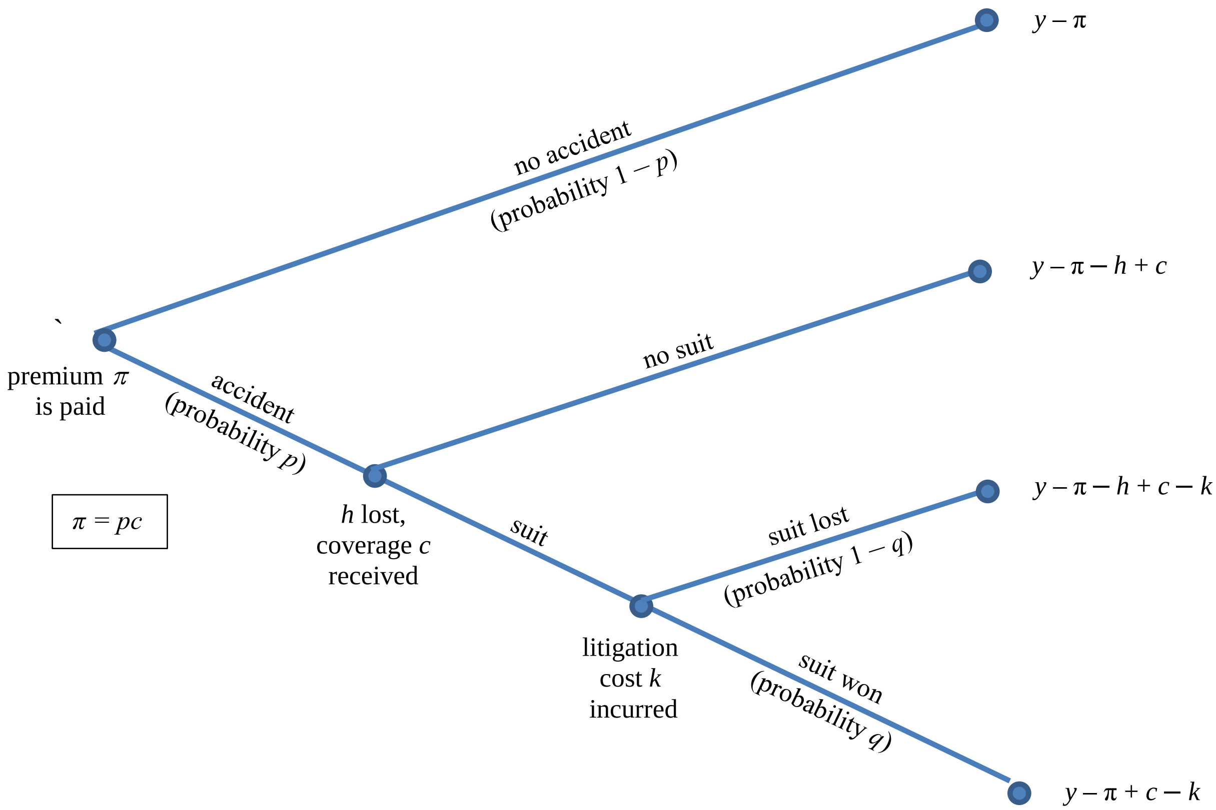

Figure 1

Outcomes under a pure insurance policy

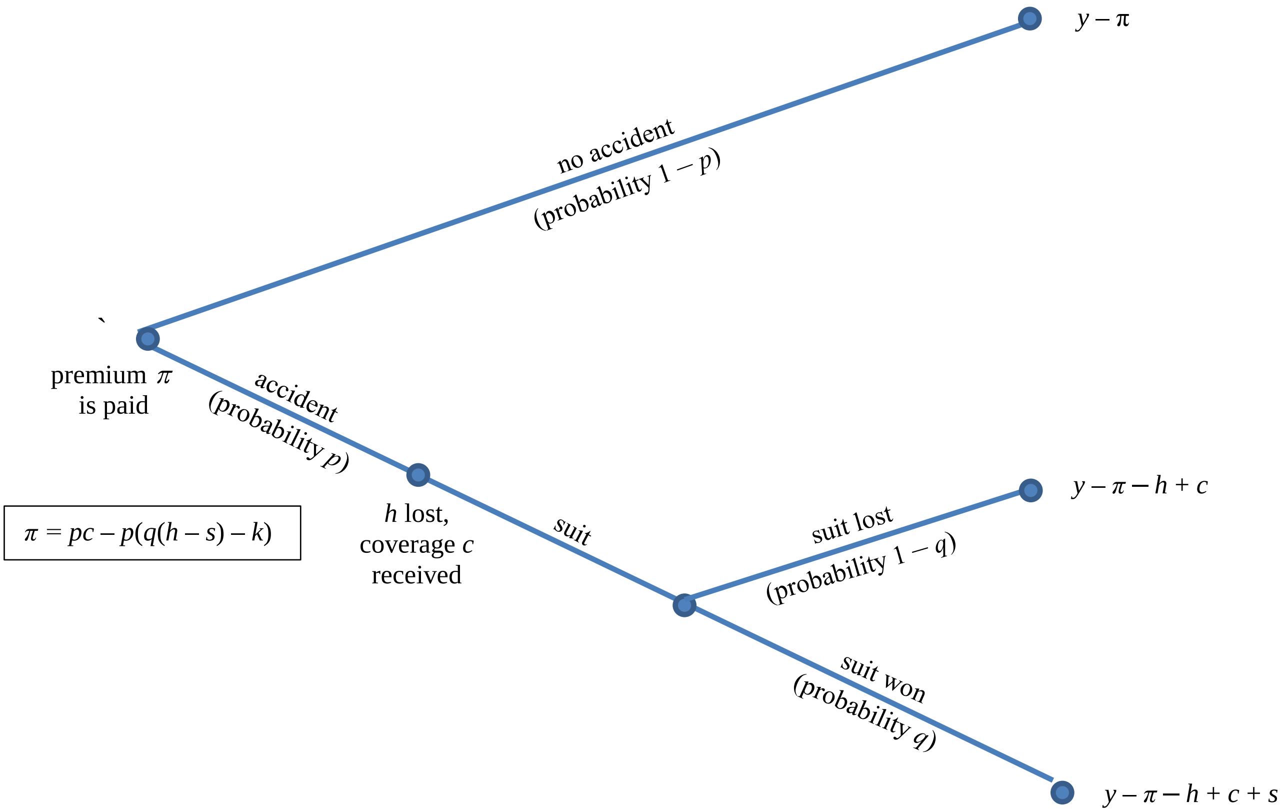

Figure 2

Outcomes under an insurance policy with a subrogation provision—when suit would be brought