11-15-2011

The Law and Economics of Liability Insurance: A Theoretical and Empirical Review

Tom Baker  
University of Pennsylvania, tombaker@law.upenn.edu

Peter Siegelman  
University of Connecticut Law School, peter.siegelman@law.uconn.edu

Follow this and additional works at: http://scholarship.law.upenn.edu/faculty_scholarship

Part of the Behavioral Economics Commons, Insurance Commons, Insurance Law Commons, Law and Economics Commons, Law and Society Commons, Risk Analysis Commons, Social Welfare Law Commons, and the Torts Commons

Recommended Citation

http://scholarship.law.upenn.edu/faculty_scholarship/350

This Article is brought to you for free and open access by Penn Law: Legal Scholarship Repository. It has been accepted for inclusion in Faculty Scholarship by an authorized administrator of Penn Law: Legal Scholarship Repository. For more information, please contact PennlawIR@law.upenn.edu.
The Law and Economics of Liability Insurance: A Theoretical and Empirical Review

Tom Baker*
and
Peter Siegelman**

*William Maul Measey Professor of Law and Health Sciences, University of Pennsylvania Law School, tombaker@law.upenn.edu.

**Roger Sherman Prof. of Law, University of Connecticut Law School, peter.siegelman@law.uconn.edu. We thank the editor, Kenneth Abraham, Sachin Pandya, Adam Scales, Daniel Schwarcz, Steven Shavell, and an anonymous referee for helpful comments and suggestions.
Abstract

We survey the theoretical and empirical literature on the law and economics of liability insurance. The canonical Shavell model predicts that, despite the presence of some \textit{ex ante} moral hazard (care-reduction by insureds), liability insurance will generally raise welfare because its risk-spreading gains will likely be larger than its adverse effects on precautionary activities. We discuss the numerous features of liability insurance contracts that are designed to reduce \textit{ex ante} moral hazard, and examine the evidence of their effects. Most studies conclude that these features work reasonably well, so that liability insurance probably does not generate substantial \textit{ex ante} moral hazard. Its effects on \textit{ex post} moral hazard (the increased tendency of victims to sue in the presence of insurance) are not as clear, however, and the welfare consequences of increased litigation are ambiguous, for reasons we explain. We discuss additional issues such as the effects of liability insurance when some defendants are judgment-proof, the problems posed by non-independence of liability risks owing to changes in legal doctrines, and the cyclical nature of liability insurance markets.
1. Introduction

Tort cases are typically denominated in terms of a victim suing an injurer. But injurers do not typically pay the damages (if any) for which they are held liable. As a result, empirical legal researchers conventionally use liability insurance premiums as a reasonably good proxy for those damages. (Baker forthcoming) The presence of insurance shapes who is sued, and for what; it shapes litigation strategy and settlement negotiations; and in some cases it even affects the underlying liability itself. (Baker 2006). Indeed, we suggest that it is not going too far to claim, echoing the great biologist Theodosius Dobzhansky, that little or nothing in Tort law makes sense except in the light of liability insurance.1 In similar fashion, Kenneth Abraham describes the relationship between tort law and liability insurance using the metaphor of a bipolar star: separate bodies that form a common gravitational field. (Abraham 2008).

Tort’s deterrence function requires that injurers internalize the costs of their behavior by paying for the injuries they cause. But liability insurance would seem to sever the link between injurers’ behavior and its financial consequences, since the party that caused the injury no longer ends up paying for it. The key insight in the theoretical economics of liability insurance is that such insurance is surprisingly likely to be compatible with robust deterrence, because insurers have various ways of coping with the moral hazard2 problem created by de-linking injurers’

---

2 Loosely speaking, moral hazard is the tendency of insurance to diminish an insured’s incentives to take care that would reduce the risk being insured against. It arises because of a classic externality: the costs of taking care fall on the insured, but the benefits of care accrue to the party who will pay for any losses, namely the insurer. For example, suppose an injurer is fully insured against liability for any accident he causes. He could eliminate any risk of a $100 accident by spending $2 on care, but without the care, the accident risk is 50 percent. If uninsured, the injurer would obviously prefer to spend $2 on care and save himself $50 (= 1/2×100) in expected liability. But with full insurance, the $2 eliminates a risk that the insured would not be responsible for, since liability is borne by his insurer. Now, the $2 is a complete waste from the insured’s perspective, even though it is in society’s interest to
financial responsibilities for injury from the costs of injury-reduction. As we explain in detail below, insurance contracts contain numerous structural features designed to limit moral hazard, and the consensus is that these features seem to work reasonably well. (On the other hand, although empirical evidence is scanty, there do appear to be some types of moral hazard that are not so well-controlled.)

Tort law is also centrally concerned with problems of compensation for victims. In this context, first-party insurance may work as an alternative to tort; society need worry less about making injurers pay victims if the latter can collect from their own insurers instead. The significance of these other compensation systems to tort law can be seen in the attention paid to them by the global reinsurance companies that reinsure liability risks. (Lahnstein 2011) Countries with more first party insurance, especially social insurance, have less need for tort law compensation. The relationship between workers compensation and tort law is similar. For example, one explanation for the size of asbestos liability losses in the U.S. as compared to other highly industrialized countries is the willingness of workers compensation and other social insurance arrangements to compensate asbestos bodily injury losses in countries such as France and Japan. (Lahnstein 2011).

2. History

Although now widely accepted in the U.S., insurance against tort (and tort-like) liability was controversial when introduced in the US in the late 19th century (McNeely (1941), Abraham 2008; Pandya 2011). The experience in Europe was similar: some opposition in the early years,
followed by acceptance and then encouragement of liability insurance in order to promote the financial responsibility of tortfeasors (Tunc 1983 at 50-5). Under Soviet law liability insurance was prohibited until recent years (Id.). The key issue, then as now, was moral hazard (Baker, 1996): many believed that allowing insurance for tort liability would lead potential tortfeasors to be less careful to prevent injuries, since they could shift the cost of any liability to their insurer.

There was often a moral or policy objection as well, founded on the belief that tort sanctions were meant to punish wrongdoers, a purpose that would be frustrated if injurers could avoid having to pay for the harms they caused. The early debates were undertaken without benefit of any real empirical evidence or theoretical framework, however, and what may ultimately have carried the day in favor of liability insurance was the absence of any organized interest with a stake in opposing it (Pandya 2011). Echoes of these earlier controversies can be heard today in more-modest debates over insurance coverage for punitive damages (Sharkey 2005), for liability arising under various employment laws (Employment Practices Liability Insurance or EPLI; e.g., Gabel, et al (2006), Mootz (1997), Gallagher (1994)), and for tort liabilities that are also crimes (Baker 2009). (See also Baker & Griffith 2010 on early objections to directors’ and officers’ liability insurance.) The reach of liability insurance has largely expanded to include these areas.

3. Justifications for the Purchase of Liability Insurance

A. Individuals

In the standard model, individuals and large corporations purchase insurance against tort liability for different reasons. Individuals are risk averse, meaning that they have a declining marginal utility of money. Insurance against liability risks, like other kinds of insurance, allows individuals to shift money from no-loss, low-marginal-utility states of the world into loss, high-
marginal-utility states of the world. When losses are shifted from the risk averse to the risk neutral (or when they are transformed from being undiversified to diversified) utility (welfare) goes up. The interested reader can refer to the Appendix for a summary of the standard analysis of the demand for and welfare effects of insurance.

Because of the law of large numbers, insurance is superior to precautionary savings for this purpose, as long as the liability risks of individuals in the pool are not too strongly correlated. In addition, individuals are sometimes required to purchase liability insurance (e.g., auto), or find it bundled with other insurance they are required to buy (e.g., homeowners insurance, which is required by lenders and which covers some liability risks).

B. Corporations

Publicly-traded corporations are understood to be owned for investment purposes by individuals with diversified portfolios and, thus, do not need insurance for risk distribution purposes. One intuition is that a fully diversified individual would own shares in both the tortfeasor and the supplier of liability insurance. Indeed, because of the downside protection from limited liability, diversified investors may actually want corporations to be risk-preferring. Yet the available evidence suggests that corporations widely purchase insurance. (Baker & Griffith 2010).

Several explanations—with varying welfare consequences—have been offered for why publicly-held corporations might nevertheless want to purchase liability insurance. One source of insurance demand is agency costs. Managers may be risk averse if they earn rents as a result of their investment firm-specific human capital, which ties their earnings to the profits of their employer (Arlen 1992: 419-20). Although shareholders can fully diversify with respect to losses of the firm, its managers can not: if the share price plummets or the firm goes bankrupt, its
managers stand to lose their rents. Indeed, managers’ contractual compensation ostensibly requires them to bear some risks in order to motivate them to avoid shirking. This then gives them an incentive to purchase insurance that reduces the variability of corporate profits, and hence of their own remuneration, even though this is not in the interests of shareholders. (Baker & Griffith 2010 provide qualitative evidence in support of this explanation).

More benign explanations for corporate liability insurance have also been offered, however. Mayers and Smith identify institutional reasons that would lead even a publicly traded corporation to benefit from shifting money from no-loss to loss states of the world and, thus, to act as if it faced a declining marginal utility of money: increasing marginal income tax rates, bankruptcy costs, and what they called the “underinvestment” problem (which relates to different incentives for equity and debt investors in certain situations) (Mayers & Smith 1982, 1987; MacMinn & Garven 2000). Froot, Scharfstein and Stein (1993) suggest that the higher costs of external capital (as compared to internal capital) constitute an additional institutional justification for corporate insurance and other forms of hedging. In addition, corporations may purchase insurance because it is bundled with other services that they value, such as loss prevention and claims management (Mayers & Smith 1982; Skogh 1989). Having the same firm sell loss control advice and insurance makes loss-control recommendations more credible, because the entity making the recommendation is also responsible for paying any losses that might occur. Insurance companies can thus more fully “bond” their loss prevention and claims management services than can separate loss-control service providers (Cohen 1997). Of course, corporations might also demand insurance because their regulators require it, or because their counterparties do (Goldberg, 2009).

Finally, privately held corporations and non-profit corporations have all of these same
reasons to purchase insurance (except that non-profits are not taxed and, to the extent that privately held corporations do not separate ownership from control, they may not face the same agency costs). In addition, privately held corporations would seem more likely to have non-diversified owners, and nonprofits, by definition, are not owned for investment purposes and, thus, are unlikely to be risk neutral (Arlen, 1992: 419).

3. The Classic Model of Liability and Insurance

The canonical formal analysis of the moral hazard effects of liability insurance is due to Shavell (1982, and for a less-formal treatment, 2004); Calabresi (1970) earlier offered significant insights. Shavell begins with the fact that insurance is socially valuable to the extent that parties are risk-averse, because bearing risk is costly. On the other hand, liability insurance may reduce social welfare to the extent that insurance enables risk-imposers to avoid bearing the expected costs of their risky activities: this is the familiar tension between the risk-spreading benefits of insurance and its potential risk-creating consequences (via moral hazard). Shavell’s (1982, 2004) contribution is to show that this tradeoff is significantly more complicated than had been previously realized, and under certain conditions, may not exist at all; even if it does, the presence of liability insurance can still be shown to be welfare-improving relative to no insurance, at least in many instances of the canonical tort model.

A. The Canonical Model without Insurance (Risk-Neutrality)

Shavell (1982) situates insurance in the canonical model of negligence and strict liability (Brown, 1973; Shavell, 1980). To set up our later analysis of liability in the presence of

---

3 This classic economic model of accidents is discussed in Jennifer Arlen, “Economic Analysis of Liability for Accidents.” in this volume. The application of this model to accidents involving customers is discussed in Andrew Daughety and Jennifer Reinganum, Economic Analysis of Products Liability: Theory and in Jennifer Arlen, Economic Analysis of Medical Malpractice Liability: Theory Grounded in Evidence, also in this volume.
insurance, we briefly review that model here, under the standard initial assumption that both injurers and victims are risk-neutral, and thus do not demand insurance. Both injurers and victims are assumed to be able to take care to prevent an accident from occurring, so we are in a two-party accident world. Care is costly to the party taking it, and thus, absent liability, injurers will not take care, since doing so reduces the probability of an accident for which they do not have to pay, and thus only benefits victims. Without insurance, victims will take the optimal amount of care, however, because they bear both the cost of care and the expected benefits (reduction in harm to themselves).

Introducing strict liability (and assuming that litigation is costless and that damages are set equal to the actual harm suffered by victims) reverses these results. (Of course, if the rule is strict liability with a defense of contributory negligence, then victims will have some incentive to take care.) Now, injurers will have the appropriate incentives to take care, because they bear the costs of any precautions taken, but also reap the benefits of care, in the form of lower expected liability to victims. On the other hand, victims are essentially fully insured because if an accident occurs, the injurer will be strictly liable for any harm that results. That leaves victims with no reason to take care, since it is costly to do so, and the benefits (a lower likelihood of an accident) accrue solely to (strictly-liable) injurers.

A key insight of the Shavell model is that injurers have more than one way of reducing accident risks: they can incur costs of precaution—“care” (such as driving more slowly) — and they can undertake an activity less often (e.g., by driving less-frequently). Under a strict liability regime, injurers will be led to consider all possible ways of reducing accident risks, because injurers will always be liable for any harms that materialize. If it is cost-effective to reduce liability by driving less-often (in addition to more-carefully), a strict liability regime will give
injurers reason to do so. In short, strict liability leads injurers to optimize on both the activity level (frequency) margin and the care/precaution margin in managing liability risk (Shavell, 1980).

A negligence regime works somewhat differently. The law defines a standard of care, compliance with which guarantees that an injurer is not negligent, and therefore not liable for any accident that may occur. An injurer taking less than the standard of care is negligent and thus liable for any accidents that result, as well as paying the costs of whatever care, if any, it has taken. Injurers will never take more than the standard of care, because doing so only reduces the probability of an accident that would not be the injurer’s fault to begin with, and hence provides benefit only to the victim. Injurers will not take less than the standard of care, either, because on standard assumptions, they will find it cheaper to comply with the standard, since doing so eliminates all expected liability, and leaves them with only the costs of care itself.4 Moreover, this model assumes that injurers can always comply because they always know when they are negligent and have full control over whether they take care. Thus, the model predicts that injurers will never be negligent as long as courts set due care equal to optimal care (Shavell, 1980), damages are set at optimal levels (Cooter, 1984; Arlen, 1985), and courts do not err in determining whether injurers and victims are negligent.5 Accordingly, in a perfect world, for

---

4 This result depends on the additional assumption that injurers can always comply because they always know when they are negligent and have full control over whether they take care. For a discussion of optimal accident law when injurers are organizations, and thus have less than perfect control of their actions see Reinier Kraakman, “Economic Analysis of Vicarious Liability,” in this volume. For an analysis of optimal accidental law when injurers may err accidentally see Jennifer Arlen, “Economic Analysis of Liability for Accidents” in this volume.

5 More refined versions of the model do predict injurer negligence under some circumstances. For example, if the standard of care is set very high, injurers may find it cheaper to take the cost-minimizing level of care—even though that is below the standard—and pay for those accidents that result. Injurers may also be negligent because of errors they make in estimating the standard of care (Arlen & MacLeod, 2005), or mistakes by courts in applying the standard to injurer behavior, among many other reasons (Cooter, 1984; Calfee & Craswell, XXX).
compensation purpose, negligence liability in effect translates into a rule of no liability to the victim, since if an accident occurs, it will never be paid for by the injurer (who is never at fault). Victims will thus take the appropriate amount of care for the same reason they do under a regime of no-liability (for injurers): in fact, victims operate under just such a regime when injurers are non-negligent.

Significantly, negligence differs from strict liability in its ability to control injurers’ activity levels. Consider driving as an example. An injurer who always drives non-negligently (takes the standard of care) will by definition never be liable under a negligence regime. Hence, such an injurer has no liability-based reason to curtail the amount of her driving, because doing so is costly and reaps her no benefits in terms of lower liability. By contrast, victims do have an incentive to control activity levels under a pure negligence regime. That’s because victims will always bear the cost of any injuries that occur (since injurers are non-negligent). Thus, victims will want to consider all options (crossing the street more carefully, or crossing less often) to reduce the risk of injury (Shavell, 1980).

B. The Canonical Model with Insurance

Table 1 summarizes the main analysis of Shavell (1982) when we assume that both injurers and victims are risk-averse, rather than risk-neutral, and hence have a demand for full insurance that is fairly priced.6 The introduction of risk-aversion makes the social decision problem more complicated. As before, we want to provide both injurers and victims with incentives to take the optimal amount of care, and to engage optimally in risky activities. In

---

6 As discussed above, corporations may demand insurance for many reasons other than risk aversion. The benign conclusions of the Shavell analysis would presumably change if the corporate demand for insurance derives from, for example, managerial exploitation of agency costs—at shareholders’ expense—rather than risk aversion. We know of no systematic theoretical inquiry into this issue, however.
addition, however, we want to spread or allocate risks optimally between parties. Bearing risk is costly for a risk-averse party, and these costs need to be factored into the social decision rule in addition to the costs of accidents themselves and the costs of accident prevention.7

Table 1 about here

1. Strict liability

Consider first a rule of strict liability, as shown in Column 1 of Table 1.8 Under this regime, victims are already in effect fully insured by injurers (who are obliged to pay for whatever accidents occur), and hence have no demand for insurance. In the extreme case, victims bear no risk and take no precautions, so the presence of insurance does not alter their behavior at all.

On the injurer’s side, strict liability causes injurers to bear the full expected cost of accidents. Risk averse injurers thus will want to insure against liability risk. Initially, it might seem that this insurance would fully undermine the incentive effects of the tort system because fully insured injurers have no incentive to invest in avoiding accidents. Yet this generally is not the case for two reasons.

First, insured injurers will retain an incentive to invest in care if liability insurers can observe their care and activity levels and price insurance accordingly. In addition, injurers will retain strong incentives to invest in care if liability insurance is “experience-rated,” in that

7 The analysis here focuses exclusively on injuries that involve replaceable commodities—i.e., things that can be translated into monetary values and thus represent pure losses to wealth. Injuries to irreplaceable commodities or entitlements (Cook & Graham 1977) may affect both the magnitude of the utility function and the marginal utility of wealth (holding wealth constant). When the injury is to a normal good, victims do not want to fully insure. Insurance demand thus looks different for losses such as death or serious permanent injury; see section 5.C below (discussing tort coverage of pain and suffering losses).

8 When insurers can observe care and activity levels, there is no informational asymmetry, and the first-best solution—to require the efficient level of care and provide full insurance—is achievable by contract.
premiums are based on the injurers’ past claims history.

Even when insurers cannot observe care or activity levels, insured injurers also will retain some incentive to invest in care (albeit weakened) to the extent they purchase insurance with a significant deductible. Partial insurance is cheaper than full insurance, because partial insurance preserves some incentive for the insured to take care, and thus generates lower premiums per dollar covered. This means that, even allowing for risk aversion, injurers will likely prefer only partial coverage.\(^9\) Partial coverage does diminish injurers’ incentives to take care, however, so there is some moral hazard effect. Shavell (1982) nicely demonstrates that the introduction of liability insurance in this context will nevertheless raise welfare (as opposed to a world of no-insurance), even though the equilibrium is only second-best optimal. The reason is that injurers obviously value the reduction in risk from partial coverage by more than the cost of such coverage (since they bought the insurance). Victims are fully insured by strict liability, regardless of whether the injurer has insurance or pays for losses itself, so victim welfare is unaffected by the introduction of insurance. And insurers are willing to sell fairly-priced insurance to injurers. Thus, there is no one made worse-off by the introduction of insurance, and injurers’ gain from bearing less risk.

2. Negligence

Now consider the operation of a negligence rule, and assume for the moment that neither courts nor injurers make any errors in assessing or applying the standard of care\(^10\) (Column 2a of Table 1). In this situation it will (almost) always be cheapest for injurers to comply with the standard of care, since this affords them complete immunity from liability. That in turn implies

\(^9\) This result might not hold if care is very unproductive and injurers are very risk-averse.

\(^10\) In other words, the standard is applied perfectly, both by injurers in deciding how much care to take ex ante, and by courts in comparing injurer conduct to the standard of care ex post.
that compliant injurers face no liability risk, and thus demand no insurance.

In this context, liability insurance does not decrease incentives to take care because injurers will not buy insurance in order to reduce their care levels below the standard of care. In this simple model, injurers face no risk of liability if they take due care, and thus insurers know that any injurer who seeks liability insurance plans to be negligent. Insurers will thus rationally price insurance assuming that those who buy it take no care whatsoever. The premium for such insurance would be very high—higher than the cost to the injurer of taking optimal (due) care.\textsuperscript{11} Indeed, since complying with the standard care is cheaper for injurers than not (because injurers will be liable for the resulting accidents), injurers will find it cheaper to comply with the standard of care than to buy insurance and slack off on care.

Knowing all this, victims will predict that they will bear the cost of any injuries they suffer, since these will be caused by a non-negligent injurer. Victims will thus demand (first-party) insurance to cover the accident risks they expect to bear.\textsuperscript{12} Full insurance would lead victims to take no care at all, and will thus be priced disadvantageously relative to partial insurance. There is thus some moral hazard that takes place on the part of victims, but despite this, victims can be shown to gain more from risk-spreading than society loses due to moral hazard, so the introduction of insurance is welfare-enhancing. (Shavell 1982) Shavell’s model also encompasses settings where the negligence standard is not applied flawlessly, either because injurers misperceive the standard of care or because courts misapply it, either of which can lead

\textsuperscript{11} If due care equals optimal care when we know that due care minimizes the total expected cost of taking care plus expected accident costs. Given this, we know that it is cheaper to take due care and pay the expected cost of accidents than to take zero care and pay a premium equal to the expected costs of accidents when injurers take zero care. Accordingly, injurers certainly are better off taking due care and not paying any accident costs than to take no care and purchase correctly priced insurance.

\textsuperscript{12} Injurers taking the standard of care are not liable, so their victims cannot expect compensation from them or their insurers. Victims’ risk of being injured by a non-negligent injurer creates the demand for first-party insurance on the part of victims.
injurers to be held negligent. In this situation, compliance with the standard of care does not immunize an injurer from liability, so injurers will demand insurance to cover this risk. The insurance will optimally be designed to protect against liability for negligence due to errors or lapses, so injurers will still find it cost-effective to take the standard of care and avoid expensive policies that would permit them to take no care. In this context, insurance improves welfare by risk-spreading, without creating moral hazard; it may also be welfare-enhancing because its presence can mitigate over-compliance that would otherwise result (in some models) from uncertainty over the standard of care (Calfee and Craswell, 1984).

Injurers might also be negligent because they can’t perfectly control their own behavior. Such imperfections can occur because exercising due care sometimes requires ongoing attention by the injurer (as opposed to “durable” precautions such as fire extinguishers that need only be installed once), and people’s attention sometimes wanders (Grady (1988), Cooter (1991)). Injurers also can be negligent because they do not have sufficient information to know what course of conduct satisfies the due care standard (as in medical malpractice) (Arlen and MacCleod, 2005). Lapses in control can also occur because of agency costs, for example when a firm is unable to perfectly monitor the behavior of employees who otherwise face insufficient incentives to invest in care (Kornhauser, 1982; Arlen and MacCleod, 2005). Under these circumstances, injurers will sometimes be negligent, even when they strive (optimally) not to be, providing further reason to conclude that there are risk-spreading benefits from liability insurance.

---

13 An example might be the restriction of pollution liability coverage to “sudden or accidental” events, which was drafted for inclusion in standard Commercial General Liability policies in the early 1970s. (Abraham 1988).
An important lacuna in the Shavell model is the assumption that all victims sue, and are fully compensated for their injuries. If this is not the case, and liability insurances creates some moral hazard, it is possible that victims might be worse off, since there will be more uncompensated injuries. We discuss this problem further in the next section.

4. Moral Hazard in Liability Insurance

A. Contract Design

As the preceding discussion implies, moral hazard has been perhaps the most significant theoretical concern animating law and economic research on liability insurance. The canonical model both relies on, and makes predictions about, the design of insurance contracts for liability risk that mitigate the potential moral hazard of liability insurance. Shavell (1982, 2004), for example, concludes that even when care levels cannot be observed directly, moral hazard can be significantly controlled through the design and implementation of insurance contracts. In large part, Shavell’s model either assumes or predicts that liability insurers will control moral hazard sufficiently so that liability insurance will be welfare enhancing. Whether and when this assumption/prediction is borne out in practice is a very important and not yet adequately answered empirical question.

In theory, the list of moral hazard control devices in liability insurance is large, and many are not unique to liability insurance. Empirical studies of their efficacy are much more limited, however, in part because data on insurance policy design and insured behavior are often proprietary or difficult to obtain. Institutional research on moral hazard control in insurance dates to Heimer (1985), though of course insurance companies have engaged in applied research for centuries (Baker 1996). Table 2 lists some of the leading studies on the ability of liability insurance contracting to limit moral hazard, organized according to the moral hazard control
technique investigated. We distinguish between \textit{ex ante} moral hazard (reduction in care that would prevent liability), \textit{ex post} moral hazard (reduction in care that would manage the extent of an asserted liability), and insurer moral hazard (shirking in contractual responsibilities). These last two categories are absent from the canonical tort/insurance model, but nevertheless play a significant role in defining the relationship between tort and insurance. Insurers employ a variety of different mechanisms to reduce moral hazard by both the insured and the insurer. These can be divided into six categories.

1. \textit{Underwriting}

Underwriting is the process of selecting among insurance applicants and assigning them to risk categories according to, inter alia, their propensity to engage in loss-prevention behavior and honesty in relation to claims practices. Underwriting has the potential to reduce both \textit{ex ante} and \textit{ex post} moral hazard.

2. \textit{Experience-rating}

Experience-rating and non-renewal: the process of adjusting the premium at renewal based on the prior claims experience and refusing to renew policyholders with the worst claims experience. Experience-rating has the potential to reduce both \textit{ex ante} and \textit{ex post} moral hazard (Robinson & Zheng, 2010).

3. \textit{Coverage design}

Insurers can design insurance coverage to reduce moral hazard by both the insured and the insurer. First, insurers can use exclusions and sublimits to eliminate or reduce the coverage available for liability losses that pose a high degree of moral hazard, most commonly because of the more substantial control that the policyholder has over the occurrence of the loss. These contract features have the potential to decrease \textit{ex ante} moral hazard. They may limit \textit{ex post}...
moral hazard as well, via provisions that require insureds to cooperate with the insurer in any lawsuit. Second, insurers also can reduce the insured’s moral hazard by requiring the insured to bear some of the cost of the accident through cost sharing and partial coverage, which can be accomplished through deductibles, coinsurance, and limits on coverage that are less than the total potential liability exposure. All of these features keep insureds’ “skin in the game,” and thus have the potential to reduce both ex ante and ex post moral hazard (Wang et al, 2010). While “aggregate deductibles” (which apply to all losses over the life of the policy) are optimal in theory (Arrow, 1963), they are rarely observed in practice, for reasons explained in Cohen (2006). Finally, insurers can reduce insurer moral hazard through the common practice of bundling defense coverage (paying the defense lawyer) with indemnity coverage (paying for settlements and judgments) and giving the insurer control (greater or lesser depending on the type of policy and circumstances) over both. Bundling has the potential to reduce insurer moral hazard, while giving the insurer control has the potential to reduce ex post moral hazard (although at the cost of creating insurer-side moral hazard).

4. Loss control

Insurers also can reduce insureds’ moral hazard by bundling loss control services with insurance. These loss control services can involve mandatory conditions or simply loss prevention advice, which should be reliable because it is bonded by the insurer’s promise to pay for any claims that result (Cohen 1997-98). Conventionally, the term “loss control” is used for activities that reduce loss or moral hazard ex ante, rather than manage the extent of the loss, ex post, but it is possible to understand defense and settlement control provisions in insurance
policies as ex post loss or moral hazard control arrangements.¹⁴

5. Ex Post auditing

Insurers can enhance the effectiveness of the moral hazard control measures falling into categories 1-4 by employing ex post auditing techniques to enforce policyholders’ representations or responsibilities under their contracts.

6. External control measures

Finally, insurers can benefit from some measures, external to insurance contracts, that operate to reduce moral hazard. Here we distinguish among the following three types of external regulation: (1) Moral or professional norms: these have the potential to reduce all three forms of moral hazard; (2) Administrative regulation: these have the potential to reduce all three forms of moral hazard; (3) Legal rules promoting the enforcement of insurance contracts: these are largely directed at reducing insurer moral hazard.

Table 2 about here

The studies listed in Table 2 generally conclude that moral hazard is present, but that institutional devices for controlling it work reasonably well (with the caveat that differences among liability and insurance institutions can make it difficult to extend the findings beyond the specific context investigated). The exception is Baker and Griffith (2010), who conclude, based

¹⁴ From the perspective of the insurance contract, “loss control” means anything that reduces liability. But controls that prevent losses ex ante have very different welfare implications from those that merely avoid injurer liability for those losses ex post. For example, EPLI insurers often insist on sexual harassment training for employees as a condition of coverage. Most such training probably has little or no effect on actual behavior—that is, on losses due to the occurrence of sexual harassment. (Grossman (2003 at 3) concludes that “... sexual harassment policies and procedures do not seem to have any reliably negative effect on the incidence of sexual harassment.”) But after the Supreme Court decided that employers were not vicariously liable for harassment by managers if the company had taken steps to prevent and promptly correct the problem (Farragher v. City of Boca Raton, 1998), sexual harassment training became a kind of “bureaucratic vaccine against lawsuits for harassment,” Dobbin and Kelly (2007, p. 1234). From a welfare perspective, there is a substantial difference between loss controls that actually reduce the incidence of harassment ex ante and those that simply shift the costs of harassment onto victims by forestalling injurer liability ex post.
on qualitative research, that the moral hazard created by directors’ and officers’ liability insurance may undercut the deterrence function of shareholder litigation by reducing the incentive of corporate boards to demand accurate financial reporting. Even they conclude that this type of moral hazard could be managed, however. If corporations disclosed the details of D&O insurance contract terms (including pricing), securities analysts could incorporate that information into a more-refined measure of corporate liability risk.

Moral hazard is difficult to measure, even with access to the necessary proprietary data, because of the need to distinguish between causal and selection effects (Abbring, Chiappori, Heckman & Pinquet 2003). Changes in insurance contract design or other techniques that are understood to control or exacerbate moral hazard produce changes in claims data through selection effects (i.e. through the mix of people who are insured) in addition to, or perhaps rather than, moral hazard (Id.; see also Cohen and Siegelman (2010) and sources cited therein for discussion of attempts to distinguish adverse selection from moral hazard). In recent work, Abbring, Chiappori and Zavadil (2008) review the prior literature and use advanced econometric techniques and Dutch automobile insurance data to separate these effects. They find both ex ante and ex post moral hazard effects from the unique experience rating system in Dutch automobile insurance, which contains features that facilitate the econometric separation of these effects.

None of the econometric studies listed in Table 2 employ techniques comparable to Abbring et al and thus, judged from the perspective of the state of the art in econometric analysis, even the findings from the other econometric studies must be regarded as provisional, at least in detail. For that reason among others, the conceptual line between “generalizable” econometric findings and “nongeneralizable” qualitative findings is less hard and fast than commonly believed. Theoretically and institutionally informed qualitative research also produces
useful empirical insights, especially but not only when the data needed to conduct econometric research are proprietary. Among other benefits, qualitative research may identify questions for subsequent econometric analysis. For example, Baker’s (2002) qualitative research describing the legal rules and professional norms that limit plaintiffs’ ability to collect “blood money” from defendants—thereby reducing the deterrence effect, if any, of liability—prompted efforts to test this hypothesis using a unique public dataset of liability claims information (Zeiler et al 2007; Hyman et al forth.; both studies support the hypothesis).

B. Ex Post Moral Hazard

The standard tort/insurance model assumes that moral hazard only occurs when insured injurers or victims take less care to prevent losses because of the presence of insurance. Insurance may lead not only to reduced care ex ante, however. It may also lead to increased claiming ex post. That might be a good thing from an efficiency perspective, if there would be too few meritorious claims in the absence of insurance (see, e.g., Andrews et al (1997) (finding that only 7% of all patients injured during their stay at a Chicago hospital made a claim based on their injury). But insurance could also generate an increase in non-meritorious claims. In the most extreme case, this might amount to outright fraud, on an organized scale (e.g., In Re: Silica Products Liability Litigation (finding systematic fraud involving 10,000 plaintiffs and sanctioning the lawyers who brought the suits in the amount of $825,000)), but more-ambiguous motivations are also possible.15

Estimates of the amount of fraud in liability insurance are difficult to come by, for at least three reasons. First, detecting fraud is difficult, and conclusively proving it even more so.

15 For an economic analysis frivolous litigation see Abraham Wickelgren, Law and Economics of Settlement, this volume
Relatedly, one wants to know not just the amount of detected fraud, but the total volume of fraud, including those cases that are not detected, which requires making difficult inferences. Finally, insurers, who control the data, may have an interest in overstating the amount of fraud. Carron & Dionne’s careful study (1997) suggests a best-guess estimate of 9.5 percent of all auto claims in Quebec involve fraud, although with a wide range of uncertainty. These authors also deploy a clever methodology that allows them to estimate undetected fraud, and which suggests that insurers are not especially good at catching fraud. Tennyson (2008) summarizes the conventional wisdom that roughly 10 percent of claims involve fraud of some kind (including over-stated damages), but goes on to suggest that “this statistic appears to be more folk wisdom than fact (at 1184) (See also Weisberg and Derrig (1991)).

What is missing from the literature is a rigorous attempt to integrate estimates of the extent and cost of fraudulent claims with the risk-spreading benefits of insurance to arrive at a net estimate of the welfare gains from insurance. A few authors (notably D’Arcy (1994), Syverud (1993-94)) have suggested that there might be a down-side to liability insurance that is not visible in the standard model, but without rigorously developing the ex post moral hazard argument. For example, Syverud suggests that liability insurance in effect creates its own demand, by encouraging litigation that would not be worthwhile in the absence of the deep-pocket insurance provides.

Table 3 provides a hypothetical example of the complex welfare calculations that might arise in the presence of ex ante and ex post moral hazard, given that there are victims who do not sue and non-victims who do. Suppose that there is initially no insurance for injurers, and that there are 100 actual injuries. Twenty five lawsuits are filed, of which 24 are “legitimate” or valid and 1 is by a non-injured party, merely seeking to extract a settlement from some
defendant. Introducing liability insurance results in a relatively modest amount of (ex ante) moral hazard on the part of injurers. The reduction in care leads to ten additional injuries, and eight additional lawsuits, of which six are valid and two are not. Sorting out the welfare implications of all this is extraordinarily complicated. First, two additional victims do not receive compensation, while 6 additional victims who would not otherwise have sued now do so. On net, therefore, the presence of insurance has beneficially raised the litigation rate and led to greater compensation and more deterrence, although some victims and injurers are worse off than before. It is thus important to distinguish between “good” and “bad” ex post moral hazard. The presence of insurance triples the number of “invalid” lawsuits (from 1 to 3). If the social cost of these invalid suits is sufficiently high (or of course if the increase is larger than in our example), the welfare costs could outweigh the benefits described above.

Table 3 about here

C. Moral Hazard in Workers Compensation Insurance

Historically, workers’ compensation (WC) represented the first large-scale attempt to substitute insurance for tort liability. WC has generated an extensive literature that is too large to discuss here in depth. The same problems of moral hazard that are central in other liability insurance relationships also appear in workers’ compensation, but greater availability of data has led to a larger volume of studies. Fortin and Lanoie (2000) survey the empirical literature on incentive effects of workers’ compensation, and conclude that the number (and also the duration) of workers’ compensation claims responds positively to the generosity of benefits (with an elasticity of about 0.6) across a wide variety of studies (at 438). They point out, however, that the studies do not allow one to distinguish between “legitimate changes in injury rates following increases in WC coverage (the reporting effect), and changes that could reflect an abusive use of
the system” (at 439), leaving the question of moral hazard open. Their bottom line conclusion is that the long-term rise in U.S. workers’ compensation costs as a percent of payroll, accompanied by a steady fall in the rate of workplace fatalities (which are “presumably a good proxy for job safety” at 453) is “worrying,” because it suggests an increase in unjustified claiming or ex post moral hazard. Yet, the aggregate costs of the health care component of workers compensation now exceed those of the income replacement component. Moreover, there is some evidence that medical care providers have shifted costs to the workers compensation system through changes in diagnosis patterns that respond to more generous reimbursements by workers compensation insurance providers than managed care organizations (Butler et al 1997). Thus, the workers’ compensation moral hazard story is a complicated one that involves much more than worker claiming patterns (Baker 1996). Confirming this conclusion, Guo & Burton (2010) find that “much of the substantial decline in actual benefits in the 1990s was due to changes in state compensability rules and administrative stringency,” rather than to worker behavior.16

5. Additional Economic Insights on Liability and Insurance

Insurance has been the site of much productive economic theorizing (e.g., Arrow 1963, Rothschild & Stiglitz, 1981). Much of that work relates to insurance generally, not specifically to liability insurance or to the relationship between liability and insurance and, thus, will not be further addressed in this review. The literature specifically related to liability and insurance has focused on the following issues.

---

16 The apparent contradiction between Fortin and Lanoie’s report of a long term rise in workers compensation costs as a percentage of payroll and Guo and Burton’s report of a decline in actual benefits in the 1990s results from differences in the data to which they had access. Guo and Burton had ten more years of data, which showed that the long term increase in workers compensation costs reported by Fortin and Lanoie stopped in the 1990s.
A. Judgment-Proof Problems

It has long been recognized that injurers can cause harm that substantially exceeds their wealth or ability to pay. (Sinn, 1982; Shavell, 1986; Gilles, 2006; Harrington and Danzon, 2000) Absent insurance, in this circumstance, victims will be inadequately compensated by any tort judgment, since injurers are unable to pay for the full extent of the harms they cause. There are efficiency consequences as well. Judgment-proof injurers will tend to make riskier decisions: Since there is effectively no liability for harms that exceed an injurers’ ability to pay, such losses are externalized to victims. As stressed by Sinn (1982), the existence of a minimum threshold for utility (e.g., via shielding of some assets in bankruptcy or through a social welfare system that provides a basic guarantee for all citizens) also reduces the demand for liability insurance by introducing non-concavities into the injurer’s utility function. (Intuitively, the judgment-proof insured may have higher expected utility from going uninsured than from buying insurance because fairly-priced insurance would have to cover and charge for losses that the insured would never have to pay if uninsured).

Smith and Wright (1992) apply these ideas to the auto insurance market, in which liability insurance is sold as a package with uninsured motorist coverage (insurance that covers first party losses in circumstances in which an at-fault driver is uninsured). They suggest that poor and/or judgment-proof drivers will have little or no demand for liability insurance. Thus, in local markets where there are large numbers of such drivers, those who do wish to purchase insurance have to pay very high premia to cover the costs of harms imposed on them by the uninsured. This in turn makes insurance even less attractive to the non-judgment proof, and can lead to multiple equilibria, in some of which, insurance is very expensive and relatively few drivers are insured. Smith and Wright’s empirical evidence on insurance costs and poverty rates
across US cities supports this conclusion.

Shavell (1986) discusses the menu of options for dealing with the judgment-proof problem, one of which is to mandate insurance purchase. Paradoxically, another option is to ban coverage altogether (Shavell, 2004, p. 278); this can be superior to mandating coverage if insurers cannot observe care levels, because injurers might be motivated to take more care than they would with insurance (even though they would take less care than would be socially optimal). Most states have some modest insurance requirements for automobile use, although these are frequently under-enforced and typically set at much lower levels of coverage than the maximum liability a driver can cause. (For details, see Baker (2008).) In the Smith and Wright (1992) model, mandating insurance (or enforcing existing mandates) can lower premiums and raise welfare. Another (complimentary) alternative is to simply regulate the conduct of the risky activity directly (e.g., nuclear power).

B. The Problem of Correlated (Socio-Legal) Risk

Danzon & Harrington (2000) point out that liability insurance will often entail non-independent risks. This dependence occurs because changes in legal precedents, institutions, or norms can influence many claims simultaneously and in the same direction. For example, a single Supreme Court decision about eligibility of asbestos victims for medical monitoring expenses (*Metro North Commuter Railroad v. Buckley* (1997)) would presumably control insurers’ liability in every one of the thousands of pending asbestos cases. Similarly, a single state supreme court decision clarifying how liability insurance contracts apply to asbestos cases could also affect how insurers’ share liability for those cases (e.g., *Owens-Illinois, Inc. v. United Ins. Co.*, 650 A.2d 974 (N.J. 1994)). While the lack of independence among risks does not make insurance impossible (Jaffe & Russell, 1997), it does make liability insurance more
difficult to price and, at least in theory, more expensive as result, reducing the net benefits from the risk distribution (Baker 2004).

**C. Tort Liability and Insurance vs. First Party Insurance**

Law and economic analysis has helped clarify the relationship between first party insurance (such as health insurance) and liability insurance (commonly referred to as third party insurance in this context). Where some early and mid-twentieth century tort theorists saw tort law and liability insurance as important victim compensation mechanisms, economic theorists identified the very high transaction costs of tort and liability insurance as compared to first party insurance. *(E.g. Calebresi 1970; see Abraham and Liebman (1992) for a useful effort to estimate the relative roles of tort and first party insurance as compensation for injuries.)* As a result, economic analysis does not consider victim compensation to be the primary function of tort law and liability insurance. Victims are efficient beneficiaries of tort law remedies (and therefore receive compensation through tort law) because their injuries are a good measure of the harm that could have been avoided by compliance with the standard of care, and because they have good access to information about those injuries, and generally the best incentives to bring claims against injurers. Of course, what matters for deterrence is that injurers pay, not that victims are compensated, so payments from injurers need not necessarily go to victims in order to achieve deterrence.

Calfee and Rubin (1992) review and extend the argument that, because tort law functions as form of mandatory victim-compensation insurance, the compensation that tort law provides ought to be limited to the kind of coverage victims would want from first-party insurance that they bought to cover their own injuries *(See also Epstein, 1985)*. This tends to suggest that since first party insurance does not ordinarily provide compensation for pain and suffering, the tort
system shouldn’t either. At best, this argument would apply to tort claims between contracting parties, such as in medical malpractice. Even in that context, however, voluntary limitations on tort damages are likely to reduce care. As Arlen explains, contract-based limits on damages present a collective action problem in which rational individual consumers may tend to choose a (socially inefficient) low level of tort protection. The reason for the inefficiency is that the existence of the warranty provides a signal of the manufacturer’s information that the goods sold are of high quality. But if a consumer can simply observe the presence of an optional warranty, without actually buying it, he can make the inference of high quality and avoid the cost of the warranty. Of course, if every consumer reasons similarly, the manufacturer will have no reason to offer the warranty.17

Further undermining the case for limited damages, Wicklegren (2005) and Arlen & MacLeod (2003) explain, patients who undervalue the insurance provided by liability (whether because they underestimate the risk of liability or because the transaction costs of tort increase the price of insurance through liability), will bargain out of liability. Finally, some experimental evidence (Avraham, 2005) suggests that people do actually demand coverage for pain and suffering, and Hanson and Croley (1995) point out that insureds may fail to buy first party

17 Steven Shavell points out (personal communication) that this argument depends on the waivability of warranties. There would be no collective action problem if a seller sold a good with an unwaivable warranty. In that case, every consumer would know that if he bought a widget from the seller, he and all other consumers will receive payments if the product turns out to be defective, so the seller would have an incentive to take care in its manufacture. While the Uniform Commercial Code restricts sellers ability to waive warranties (see, e.g., §§2-314 through 2-316), it does so to “protect a buyer from unexpected and unbargained language of disclaimer” (§2-316, Comment 1). The Code does not appear to prohibit sellers from making unwaivable warranties. Arlen (2010) suggests that in the medical malpractice context, at least, the only realistic way to achieve an unwaivable warranty would be through a standard form contract (because of the problem of duress in individual negotiations), which would present an adverse selection problem. The products liability context also presents a risk of adverse selection, provided that the risk of injury is a function of both product safety and victim behavior. Arlen points out (personal communication) that whether this undermines the case for contract would turn on whether consumers are better off under a pooling equilibrium with cross-subsidization from low risk to high risk, or under a separating equilibrium with low risks signaling their status by waiving liability (and getting low cost, but less safe products than they want) and high risks purchasing the warranty.
insurance for certain kinds of losses not because they don’t want such coverage, but because insurers will not supply it for fear of adverse selection or moral hazard.

Like moral hazard, adverse selection (Rothschild and Stiglitz, 1981) requires an informational asymmetry. Here, however, what is unobservable (to the insurer) is not the insured’s level of care, but rather the insured’s exogenously given level of riskiness. The theory of adverse selection predicts that when insureds know their own riskiness better than insurers do, the best risks will tend to select-out of the insurance pool, since they find insurance that is priced for the average risk to be unattractive to them. The worst risks will be delighted to buy insurance priced for the average person. As a result, insurers will end up selling policies priced for the average risk to only the worst risks, which is obviously adverse to their interests. Cohen and Siegelman (2011) provide an empirical/methodological survey of adverse selection in insurance markets.

Interestingly, the problem of adverse selection in liability insurance has attracted much less scholarly attention than that of moral hazard, although both are widely understood to plague insurance markets in general. It may be that insureds do not have superior information vis-a-vis insurers with respect to liability risk. Alternatively, perhaps the methods for controlling selection are sufficient to prevent the worst risks from pooling with better ones. An influential article by Priest (1987) blamed an adverse selection-like mechanism for the liability insurance crisis in the mid-1980s, suggesting that a variety of pro-plaintiff decisions caused unraveling in several sub-markets for liability insurance. Even if the law moved in the way Priest suggested,

18 Recent work by Arlen (2010) and Choi and Spier (2010) takes up the problem of adverse selection in medical malpractice and products liability contexts. Both papers suggest that adverse selection may justify mandatory (non-waivable) liability; both papers conclude that allowing waivers may lead to inefficient un-pooling of heterogeneous customers.
however, it is difficult to see how new legal rules could create asymmetric information favoring insureds (Ayres and Siegelman, 1989). (Absent such informational asymmetry, insurers would simply charge more for additional liability risks, and insureds would agree to pay more for coverage, with no unraveling of the market.) In sum, the scholarly literature does not seem to consider adverse selection to be a serious problem in liability insurance markets.

As compared to first party insurance, tort law and liability insurance present less risk of adverse selection (because tort law applies to everyone, and in that sense is mandatory) and less risk of ex post moral hazard (because it is harder to fake a tort than an insurance claim\textsuperscript{19}). In any event, because compensation is not the primary economic function of tort law, the debate over what level of tort “insurance” consumers are or should be willing to pay for is beside the point. Thus, for example, whether consumers do or do not purchase first party insurance for pain and suffering losses should not be relevant to the law and economic analysis of tort damages.\textsuperscript{20} From an economic perspective, the appropriate measure of damages is that which will induce compliance with the standard of care (Geistfeld 1995).\textsuperscript{21}

\textit{D. The Liability Insurance Underwriting Cycle}

\textsuperscript{19} A tort claim is harder to fabricate than a first party insurance claim because there is – in most cases – a third party, in addition to the insurance company, with a significant interest in verifying the accuracy of the claim: the defendant. In some cases, of course, defendants and plaintiffs may have a sufficiently close relationship that the defendant may be willing to “admit” to a liability that didn’t exist. For that reason, liability insurance contracts often contain provisions that exclude coverage for claims brought by people with too close a relationship with the insured party. Examples include the family member exclusion in personal liability insurance policies and the insured versus insured exclusion in directors’ and officers’ insurance policies.

\textsuperscript{20} It is worth noting that some first party insurance for pain and suffering damages is widely purchased: uninsured and underinsured motorists coverage. While a certain level of coverage typically is mandated, many people voluntarily purchased much higher levels of coverage. Moreover, because tort-based compensation poses less adverse selection and moral hazard (at least for victims) than first party insurance, tort based compensation can be more complete than voluntary first party insurance.

\textsuperscript{21} Note that Arlen & MacLeod (2005) would modify this in situations in which injurers face a significant risk of accidental negligence (such as medical malpractice) because it is costly for them to obtain the information needed to know what action satisfies the standard of care and where optimal investment in this information still leaves them with a positive probability of error. In this case, tort damages must both induce informed injurers to want to comply with the standard of care but also ensure that injurers will want to invest optimally in information.
The liability insurance crisis of the mid 1980s prompted extensive research that acquainted law and economics scholars with the insurance underwriting cycle, an insurance industry business cycle characterized by relatively long periods of stable nominal (declining in real) insurance prices and periodic short periods of rapidly increasing insurance prices and contractions in supply (Harrington 2004). In a prominent and widely cited early contribution, Priest (1987) asserted that the crisis was caused by the shift to strict products liability, with resulting adverse selection consequences attributable to manufacturers’ private information regarding the safety of their products. Schwartz (1992) subsequently concluded that product liability law did not change so fundamentally during the years leading up to the crisis and, as Harrington & Danzon (2000) report, no systematic empirical evidence has been produced to support Priest’s adverse selection hypothesis. The academic literature has settled on the view that the mid 1980s liability insurance crisis was an extreme dip in the longstanding underwriting cycle in property casualty insurance (e.g. Winter 1989), perhaps exacerbated by a mid 1980s change in taxation rules governing the reserves held by property casualty insurance companies (Logue 1995).

The liability insurance crisis of the early 2000s prompted another round of research (Harrington et al 2008; Baker 2005). This research supported the emerging consensus that (a) competition leads to excessive price cutting in liability insurance markets (perhaps because of weak incentives for solvency) that (b) gradually eliminates insurance capital until (c) a tipping point is reached (perhaps because of an external shock), at which point (d) insurers sharply raise prices and loss expectations, creating an extreme short-term capital shortage that allows insurers to earn large profits, prompting, with a delay, (e) external capital to flow into the insurance market (Gron 1994, Gron and Winston 2001) that, with another delay, leads back to (a) another
round of excessive price cutting, and so forth. (Harrington et al 2008) (See Fitzpatrick 2004 for a behavioral explanation from an insurance industry insider.) So understood, the large increase in expected losses that accompany the crisis stage of the cycle is an effect of the crisis, not a cause (Baker 2005). A very recent comparison of U.S. liability and property insurance premiums from 1958 to 2008 shows that aggregate real premiums for both kinds of insurance have grown at approximately the same rate over this period as real growth in U.S. GDP (Baker 2011). The pattern of growth differs, however, with liability insurance premiums exhibiting much greater growth during the peaks of the insurance underwriting cycle, when media coverage of liability insurance also peaks, suggesting that the popular perception that liability costs are growing much more rapidly than the economy may be the result of generalizations from a biased set of observations (Id.)

E. Liability Insurance Data as a Window on Tort Claiming and Litigation

Empirical law and economics researchers have made widespread use of liability insurance data to study the operation of the tort system. Relying on Texas Department of Insurance medical malpractice data, one productive research team has documented that plaintiffs rarely receive the amount that juries award, settlements rarely exceed liability insurance policy limits, and the dollar value of medical malpractice claiming has declined in real terms since the late 1990s (E.g. Black et al 2005; Hyman et al 2007; Zeiler et al 2007; see also Viscusi et al 2007 for similar research using Florida data). Carrol & Abrahamse (2005) used automobile insurance data from the Insurance Research Council to measure the impact of different tort law rules on the rate of claims arising out of soft tissue injuries, finding strong evidence of the “bad” kind of ex post moral hazard discussed above. Hersch and Viscusi (2007) used the Texas Department of Insurance commercial liability insurance settlement database to provide useful descriptive
statistics about commercial litigation, which has not received the same attention in the empirical literature as personal injury litigation. Born and Viscusi have used the aggregate liability insurance data filed by insurance companies with state insurance regulators to document that caps on tort damages reduce aggregate insurance losses (E.g., Born et al 2009).

Yoon has used data from individual insurance companies to measure the impact of limits on punitive damages and changes in rules regarding offer of judgment, finding that both sets of legal reforms achieve their intended results (Yoon 2001, Yoon and Baker 2006). In addition, a variety of researchers have used insurance companies’ closed claim files to conduct audits on the performance of tort litigation, largely concluding that the tort claim process “works” in the sense that payments are strongly and appropriately correlated with evidence of injury and negligence (e.g. Studdert et al 2006).

6. Conclusions

The canonical model of the interaction between insurance and tort liability is too stark and simple to be a satisfactory description of the real world, but it nevertheless offers some important insights. It suggests that the interaction of insurance with the legal rules for liability may not cause much moral hazard after all; and even when it does, it is still likely to be welfare-enhancing, because the risk-spreading gains are typically larger than the losses due to reduced investments in care or precautions. Thus, the existence of some moral hazard is not necessarily a system failure in a second-best world (where care is unobservable). The model also makes predictions about the kind of insurance contracts that we should observe: moral hazard will tend to be limited by pricing and contractual design, so as to discourage substantial deviation from the standard of care (a standard with which uninsured injurers operating under a negligence regime would arguably wish to comply in the absence of insurance). The standard model also directs
attention to the failures of ordinary tort law: when injurers can evade liability because they are judgment proof, or because causation can be difficult to prove, efficiency may require regulation of liability insurance coverage—either mandating it and banning it. The range of potential contract designs is vast, however. Greater knowledge of institutional detail and much more empirical investigation is needed to identify and understand which designs are employed in which context and why. Ideally, the results of this empirical research will be fed back into the modeling through an iterative process that improves our understanding of the dynamics behind the real world of liability and insurance.
**Table 1: Shavell’s Analysis of The Effects of Liability Insurance**

<table>
<thead>
<tr>
<th>Actor</th>
<th>Liability Regime</th>
<th>Strict Liability</th>
<th>Negligence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Victim</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Insurance Purchase? (1st Party)</td>
<td>None, already fully-insured by SL</td>
<td>Yes, b/c injurer will be non-negligent, regardless of liability insurance, so victim will bear all losses</td>
<td></td>
</tr>
<tr>
<td>Risk-Bearing?</td>
<td>None</td>
<td>Some, b/c of incomplete coverage</td>
<td></td>
</tr>
<tr>
<td>Care Incentive?</td>
<td>None</td>
<td>Some, b/c of incomplete coverage</td>
<td></td>
</tr>
<tr>
<td>Activity Incentive?</td>
<td>None</td>
<td>Some, b/c of incomplete coverage</td>
<td></td>
</tr>
<tr>
<td><strong>Injurer</strong></td>
<td></td>
<td>(a) No Errors in Negligence Rule</td>
<td>(b) Possible Errors in Negligence Rule</td>
</tr>
<tr>
<td>Insurance Purchase?</td>
<td>Probably partial, even when insurers are willing to provide full insurance</td>
<td>None</td>
<td>Yes, designed to cover their own lapses or judicial error**</td>
</tr>
<tr>
<td>Risk Bearing?</td>
<td>Sub-optimal, injurer retains some risk</td>
<td>None: injurer will take std. of care, thus never liable</td>
<td>None</td>
</tr>
<tr>
<td>Care Incentive?</td>
<td>Sub-optimal, injurer takes insufficient precaution</td>
<td>Injurer will take std. of care.</td>
<td></td>
</tr>
<tr>
<td>Activity Incentive?</td>
<td>Sub-optimal, injurer engages in activity too frequently</td>
<td>None</td>
<td>Policy may require some control over activity</td>
</tr>
<tr>
<td><strong>Overall Assessment:</strong></td>
<td>Insurance improves <strong>injurer</strong> welfare by risk-spreading, leaves <strong>victims</strong> indifferent, and covers its cost to <strong>insurers</strong> <em>via</em> fair premium. Hence, it is an unambiguous improvement over no-insurance, although still not first-best.</td>
<td>Insurance is irrelevant to injurer, because she has already ‘purchased’ full immunity from liability by taking std. of care. Victims will purchase partial insurance, which reduces risk, but dilutes incentives to take care, with overall net gains to victims and no effect on injurers or insurers.</td>
<td>Insurance helps inj’s avoid liability for own lapses and/or judicial error; will be designed and priced to maintain inj’s’ incentives to take std of care. Inj’s gain from risk spreading, w/out incr. in moral hazard. Ins. may <em>raise</em> welfare by mitigating over-compliance or controlling activity level. Victims are indifferent.</td>
</tr>
</tbody>
</table>

*Assumptions (based on Shavell, 1982): All parties are risk-averse. Insurer cannot observe amount of care or activity level. The regimes are pure strict liability or negligence, with no contributory or comparative elements. Premiums are actuarially-fair. Losses are purely pecuniary. Injurers can always pay for any harm they may cause.

**Policy will not be attractively priced for an injurer wishing to be deliberately negligent.
<table>
<thead>
<tr>
<th>Technique</th>
<th>Effect</th>
<th>Literature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Underwriting: screen in advance for “good character” or other evidence of good loss control</td>
<td>Reduce ex ante moral hazard, possibly reduce ex post</td>
<td>Heimer (1985); Glenn (2000); Baker &amp; Griffith (2010)</td>
</tr>
<tr>
<td>Experience rating (&amp; non-renewal): Charge more for (or drop) those with poorer claims histories</td>
<td>Reduce ex ante and ex post moral hazard</td>
<td>Cohen (???); Derrig &amp; Tennyson (2010); Israel (2004); Robinson &amp; Zheng (2010)</td>
</tr>
<tr>
<td>Coverage design:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. Exclusions and sublimits that scale coverage to reflect injurers’ control</td>
<td>A. Reduce ex ante moral hazard</td>
<td>A. Heimer (1985); Baker (2008)</td>
</tr>
<tr>
<td>B. Bundling defense &amp; indemnity</td>
<td>B. Reduce ex post moral hazard and reduce insurer moral hazard</td>
<td>B. Danzon (1985); Sykes (1994); Baker (1998b)</td>
</tr>
<tr>
<td>C. Cost-sharing &amp; partial coverage, incl. deductibles</td>
<td>C. Reduce ex ante and ex post moral hazard</td>
<td>C. Heimer (1985); Baker &amp; Griffith (2010); Wang et al (2008)</td>
</tr>
<tr>
<td>Ex post auditing to enforce the techniques above</td>
<td>Reduce ex ante and ex post moral hazard</td>
<td>Dionne (2000); Picard (2000); Ericson &amp; Doyle (2004); Tennyson (2008); Derrig &amp; Tennyson (2008); Weisberg &amp; Derrig (1991)</td>
</tr>
<tr>
<td>Reliance on other regulators</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. Moral or professional values</td>
<td>Reduce ex ante, ex post, and insurer moral hazard</td>
<td>A. Cummins &amp; Tennyson (1996); G. Cohen (1997-98); Tennyson (2002); Baker (2001)</td>
</tr>
<tr>
<td>C. Legal rules. E.g.:</td>
<td></td>
<td>D. Choi &amp; Liang (2007)</td>
</tr>
<tr>
<td>• Duty to settle</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Damages for bad faith breach</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Waiver of coverage defense upon breach of duty to defend</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Prejudice rule for breach of conditions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Anti-fraud laws</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D. Organizational form</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3: Hypothetical Effects of Introducing Liability Insurance on Volume & Composition of Litigation, When Not All Victims Sue, and Some Non-Victims also Sue

<table>
<thead>
<tr>
<th>Description of Effect</th>
<th>No Insurance</th>
<th>Insurance</th>
<th>Increase</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Ex Ante Moral Hazard</td>
<td>100</td>
<td>110</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Ex Post Moral Hazard</td>
<td>25</td>
<td>33</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>&quot;Good&quot; Ex Post Moral Hazard</td>
<td>24</td>
<td>30</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>&quot;Bad&quot; Ex Post Moral Hazard</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>No Insurance</th>
<th>Insurance</th>
<th>Increase</th>
<th>Description of Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number Injured</td>
<td>100</td>
<td>110</td>
<td>10</td>
</tr>
<tr>
<td>Total Lawsuits</td>
<td>25</td>
<td>33</td>
<td>8</td>
</tr>
<tr>
<td>Litigation Rate</td>
<td>25%</td>
<td>30%</td>
<td>6</td>
</tr>
<tr>
<td>&quot;Valid&quot; lawsuits</td>
<td>24</td>
<td>30</td>
<td>6</td>
</tr>
<tr>
<td>&quot;Non-Valid&quot; lawsuits</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>
Figure One
Decreasing Marginal Utility of Wealth

Utility of Wealth

U(wealth)

Flat Slope = Low Marginal Utility of Wealth

Wealth

Steep Slope = High Marginal Utility of Wealth
Figure Two
Demand for Insurance

Utility of Wealth

$U(W')$

$U(W)$

$U(W-P)$

Avg. Utility from Gamble

$U_e$

$W$

$W - P$

$W'$

Wealth

$W = 2W$

$0$

Wealth

Utility of Wealth

$U(W')$

$U(W)$

$U(W-P)$

Avg. Utility from Gamble

$U_e$

$W$

$W - P$

$W'$

Wealth

$W = 2W$

$0$
Appendix
Risk Aversion and the Demand for Insurance

The demand for insurance arises from risk aversion on the part of insureds. Risk-aversion is a term of art, which has nothing to do with shunning physical risks such as snow-boarding or rock climbing. Rather, risk aversion means only that someone has a decreasing marginal utility of wealth—as their wealth increases, the gain from adding another dollar to their wealth (the slope of the utility function) gets smaller and smaller. This concept is depicted in Figure 1.

Risk averse individuals experience equal-sized monetary gains and losses asymmetrically in utility terms: not all dollars are created equal. Consider Jane, who has initial wealth $W$, as in Figure 2. If she were to experience a gain in wealth of 100%, she would move to wealth level $W'$. But notice that because of the curvature of her utility function, this 100% gain in dollar terms translates into a much smaller percentage gain in utility (well-being), as measured by the small rise in utility from $U(W)$ to $U(W')$ on the vertical axis. Even though $W'$ is twice as large as $W$, Jane’s utility goes up by a factor of much less than this. The reason is that her wealth gets converted into utility at a less- and less-favorable rate as wealth increases. The converse is also true. A drop in wealth hurts much more in utility terms than its cost in dollars, because each dollar lost is worth more and more (in terms of utility) as Jane has fewer of them.

Now suppose that Jane starts with wealth $W'$, and potentially faces liability for an injury that she might cause to someone. The injury occurs with probability one-half, and if it occurs, Jane will be liable for the entire amount of her wealth. Thus, she faces two possibilities: wealth of 0, or wealth unchanged at $W'$. Her average or expected wealth is just

$$\text{Average Wealth} = \frac{1}{2} \times 0 + \frac{1}{2} \times W' = W,$$

which lies midway between 0 and $W'$. But Jane’s average (or expected) utility is
Average Utility = \( \frac{1}{2} \times U(0) + \frac{1}{2} \times U(W') \),

which is depicted in the graph as \( U_e \), and is half-way between 0 and \( U(W') \). Notice that this is considerably less than the utility she would get from having wealth \( W \) with certainty, which is given by \( U(W) \).

This forms the basis for insurance. A risk-neutral insurance company\(^{22}\) would be willing to take on this risk if it could charge a premium that was equal to its average payout. The payout is either 0 (if the loss doesn’t occur), or \( 2W \) (if it doesn’t), each with 50% probability. In other words, the average payout is just \( W \).

Were the insurance company to charge her \( W \) to assume the risk, Jane would then have a guaranteed wealth of \( (W' - W = W) \), and a guaranteed utility of \( U(W) \).\(^{23}\) Since this is higher than the expected utility from bearing the risk herself (\( U_e \)), she would be better off with insurance than without it. In fact, she would be willing to pay somewhat more than \( W \) to be free of the risk. Consider starting at wealth \( W' \) and paying an amount \( (W + P) \) to eliminate the risk altogether: that leaves Jane with certain wealth \( (W - P) \), which nevertheless has a utility level that is higher than that associated with the gamble (\( U_e \)). Put differently, insurance is a way to move wealth from states of the world where it is relatively plentiful (and hence, where an extra dollar is worth little in utility terms) to states of the world where wealth is scarce (and where an extra dollar is very valuable in utility terms).\(^{24}\)

---

\(^{22}\) For these purposes, we assume that there are no costs to running the insurance company; this is obviously incorrect, but it makes the exposition simpler without fundamentally changing the analysis.

\(^{23}\) Jane would pay her insurer \( W \). If the risk materialized (which happens with probability \( \frac{1}{2} \)), Jane would receive \( 2W \), leaving her with wealth of \( 2W - W = W \). If the risk did not materialize (also with probability of \( \frac{1}{2} \)), Jane would receive nothing, leaving her with wealth of \( 2W - W = W \). So her wealth is \( W \) regardless of whether the risk occurs or not.


Abraham, Kenneth S. (2008), The Liability Century.


Arrow, Kenneth J. (1963), Uncertainty and the Welfare Economics of Medical Care, 53 AMER. ECON. REV. 941.


Baker, Tom (1998b) Liability Insurance Conflicts and Defense Lawyers: From Triangles to


Baker, Tom (forthcoming), Transparency through Insurance: Mandates Dominate Discretion, in TRANSPARENCY IN THE UNITED STATES CIVIL JUSTICE SYSTEM: AN EMPIRICAL ANALYSIS WITH PRESCRIPTIONS FOR REFORM (Joseph Doherty & Robert T. Reville eds., forthcoming)


Butler, Richard J., Robert P. Hartwig and Harold Gardner (1997), HMOs, Moral Hazard and Cost Shifting in Workers Compensation, 16 J. HEALTH ECON. 191


Cohen, Alma (2006), *Disadvantages of Aggregate Deductibles*, 6 BE JOURNALS IN ECONOMIC ANALYSIS & POLICY.


Cook, Philip J. and Graham, Daniel A (1977), *The Demand for Insurance and Protection: The Case of Irreplaceable Commodities*, 91 Quart. J. Econ. 143.


Substance in Sexual Harassment Law, 26 HARV. WOMEN’S L. J. 3.


McNeely, Mary Coate (1941), Illegality as a Factor in Liability Insurance, 41 COL. L. REV. 26.


Shavell, Steven (1982), *On Liability and Insurance*, 13 BELL J. ECON. 120.


