8-1984

Tort Reform and the Role of Government in Private Insurance Markets

Patricia. M. Danzon
University of Pennsylvania

Tort Reform and the Role of Government in Private Insurance Markets

Disciplines
Insurance | Other Business

TORT REFORM AND THE ROLE OF GOVERNMENT IN PRIVATE INSURANCE MARKETS

PATRICIA M. DANZON*

PUBLIC concern over liability rules in recent years has been triggered by disruption in liability insurance markets. The medical malpractice and product liability "crises" of the mid-seventies were marked by insurance premium increases of several hundred percent in a single year and complete withdrawal of commercial carriers from some areas. Although insurance is now generally available, high premium rates for some product lines remain controversial.[1] A related and growing concern is the inadequacy of existing insurance and self-insurance reserves to cover product liability claims for occupational disease.

These disruptions have led to two types of statutory response. The first, which works on the tort system directly, includes changes in underlying liability rules, redefining damages, the standard of care, and the statute of limitations. The second operates on insurance markets, through rate regulation and various subsidy mechanisms. This paper evaluates these two forms of intervention, noting that the tort system may be viewed as a system of compulsory insurance,[2] with terms of coverage determined largely by the private choices that generate court decisions. Statutory tort reform may be viewed as a collective choice to override the private choices reflected in the evolution of common law.

[1] Hearings on the Nature and Causes of the Product Liability Problem Before the Subcomm. on Consumer Protection and Finance of the House Comm. on Interstate and Foreign Commerce, 96th Cong., 1st Sess. (1979).

[2] See, for example, Walter Y. Oi, The Economics of Product Safety, 4 Bell J. Econ. 3 (1973).

I adopt the premise that the criterion for an efficient liability rule is the minimization of costs from four sources: injuries, injury prevention, risk bearing, and the overhead cost of litigation and administration. In the case of medical malpractice and product-related injuries, some transfer of liability from victim to injurer seems optimal, because of information asymmetries and contracting costs. Imposing liability on injurers promises gains in prevention, and perhaps risk allocation, that outweigh the additional overhead costs. But it remains to define the limits of the optimal transfer of liability on several margins. Here I ignore the fundamental question whether the liability rule should be negligence or strict liability and focus instead on the amount of compensation and the period in which claims may be brought.

Any rule of third-party liability provides the potential beneficiary with some degree of compulsory insurance and, like any insurance, thereby invites moral hazard on several dimensions. Here I ignore the question of reduced incentives for injury prevention by the potential victim. My concern is with incentives to file exaggerated or invalid claims. I argue that, whether because of this moral hazard or other biases, liability on both margins—size of awards and period for suit—has been pushed beyond the socially optimal level, given the resulting impact on deterrence, risk allocation, and overhead costs. Statutory limitations on damage awards and on the period for suit (statutes of limitations or repose) would improve the overall efficiency of the tort system in ways that have been neglected by courts ruling on the constitutionality of such legislation.[3] These limits would also go a long way toward resolving the perceived problems in liability insurance markets and are superior to the alternative of direct government intervention.

The paper is in two parts. Part I discusses awards. Section IA describes the theoretically optimal structure of awards for the dual purpose of compensation and deterrence. The formal model in Appendix A, following Spence,[4] shows the optimality of a two-part rule of damages: (1) a compensatory award, paid to the victim, determined by the amount of insurance he would have chosen to buy given the price of the defendant's liability insurance; and (2) a "deterrence surcharge" or fine on the defendant, paid to the state, determined by potential victims' willingness to pay for injury prevention and by the uninsured costs of suit to the defendant. Empirical evidence is then presented to provide rough dollar counterparts to these theoretical constructs. The tort norm of full compensation is

---

[3] For example, 25 A.L.R. 4th 643 (1983).

[4] A. Michael Spence, Consumer Misperceptions, Product Failure and Producer Liability, 44 Rev. Econ. Stud. 561 (1977).

shown to provide more insurance than people would voluntarily choose. While this conclusion and several parts of the underlying analysis are not new, my purpose is to pull them together and draw explicit inferences for tort reform. Sections I*E* and I*F* analyze two specific proposals. The first replaces the existing system of individually determined awards with a schedule of payments, similar in principle but not necessarily in magnitude to workers' compensation benefit schedules. The second integrates tort with other private and public sources of coverage, by providing for subrogation rights of all private coverages against the tort award and reduction of public benefits by the amount of tort recovery.

Part II analyzes the impact of long statutes of repose for latent injuries and long-lived products, when liability rules are changing and uncertain.[5] Anticipated changes may either force an intergenerational transfer or expose producers or victims to risk. Uncertain changes create sociolegal risk that is independent of the size of the risk pool and cannot be costlessly diversified through multiline insurance portfolios or stock markets. A model of the relationship between the statute of limitations, the insurance risk, and the price of insurance is developed (Appendix B). The concluding subsection argues for a shorter statute of repose, shifting compensation for long latent injuries to first-party mechanisms, in preference to special government funds.

I. SIZE OF AWARDS

Awards for catastrophic injuries account for a very large fraction of the cost of the tort system. Three percent of all medical malpractice claims (or 5 percent of paid claims) account for 50 percent of the total dollars paid out. Since the early seventies, tort awards, especially for severe injuries, have grown more rapidly than the rate of inflation.[6] In response

---

[5] Several different definitions of "statute of repose" are in use. See Francis E. McGovern, The Variety, Policy and Constitutionality of Product Liability Statutes of Repose, 30 Am. U. L. Rev. 579 (1981). I use "statute of repose" to refer to a statute that places a limit on the time in which actions may be brought under traditional statutes of limitations. This may be done either by setting an outer limit on a "discovery" provision or by setting the time at which the statute begins to run at a different point from traditional tort statutes of limitation, which begin to run only when the cause of action, including the manifestation of the injury, has accrued.

[6] For the period 1971–78 the Insurance Services Office estimates annual growth rates in average incurred cost per claim as follows: physicians and surgeons, 12.1 percent; hospitals, 18.9 percent; product liability bodily injury, 19.4 percent; automobile bodily injury, 14.1 percent. Patricia M. Danzon, The Frequency and Severity of Medical Malpractice Claims, 27 J. Law & Econ. 115 (1984). See also Mark A. Peterson & George L. Priest, The Civil Jury: Trends in Trials and Verdicts, Cook County, Illinois, 1960–1979 (R-2881-ICJ, Rand Corp. 1982).

to the liability crisis of the mid-seventies, many states enacted tort "reforms" designed to reduce awards. Of these changes, dollar caps and mandatory offset of compensation from collateral sources appear to have had a significant impact,[7] but in several states they have not survived constitutional challenge.

A. The Optimal Structure of Tort Awards

Under current rules, tort damage awards may have both a compensatory and a punitive component. Compensatory awards are intended to compensate fully for pecuniary and nonpecuniary losses. In addition to compensatory damages, punitive damages may be awarded in cases of reckless disregard.

Cook and Graham and Spence have shown that the principle of full compensation is unlikely to be optimal from the standpoint of either deterrence or compensation, at least in the case of seriously disabling injuries.[8] In Appendix A, I develop a model of the optimal structure of awards when the potential injurer (producer) and victim (consumer) are in a market relationship and the probability of injury depends only on the prevention effort of the producer.[9] The model extends that of Spence, by allowing for uninsured costs of suit to the injurer, and by recognizing the expense loading in liability insurance premiums. The effect of consumer misperception of risk is analyzed, since this provides the primary rationale for restrictions on private contracts and mandatory third-party liability in these markets.

The main implications of this model are that the optimal award depends on the type of injury, the expense load of the defendant's liability insurance, and the extent to which contractual relations between the parties transmit ex ante consumers' true willingness to pay for safety. There are two issues: compensation and deterrence. Optimal compensation depends only on the type of loss and the expense loading. The optimal compensatory award is the amount of insurance the victim would have purchased voluntarily, at the price implied by the load of the defendant's liability insurance.[10] This is intuitively obvious, since the tort system provides compulsory insurance, purchased from the defendant or ultimately from his liability insurer. The tort norm of full compensation is

---

[7] Danzon, *supra* note 6.

[8] Philip J. Cook & Daniel A. Graham, The Demand for Insurance and Protection: The Case of Irreplaceable Commodities, 91 Q. J. Econ. 143 (1977). Spence, *supra* note 4.

[9] The model can readily be extended to include victim precautions. See Samuel A. Rea, Jr., Nonpecuniary Loss and Breach of Contract, 11 J. Legal Stud. 35 (1982).

[10] This analysis ignores social benefits from private enforcement of laws.

optimal only if the loss is purely monetary and the load is zero. But for a serious injury that affects the utility of wealth, optimal compensation could be more or less than the monetary loss, depending on whether disability raises or lowers the marginal utility of wealth, which cannot be determined a priori.[11] Full compensation for pain and suffering is unlikely to be optimal, because insurance can only transfer money from the healthy to the disabled state, but money cannot replace the nonpecuniary losses of physical injury. Optimal insurance transfers dollars only to the point where the victim values the marginal dollar as much when disabled as when he is healthy. With a positive load, optimal coverage is lower.

If optimal compensation is less than full, so the victim is not indifferent to the occurrence of injury, the compensatory award alone may provide insufficient incentives for prevention. If consumers are in a market relation with producers, and correctly perceive the risk of injury ex ante and can costlessly monitor contract performance ex post, consumer willingness to pay for injury prevention is internalized to producers through markets. In that case, the compensatory award also suffices for optimal deterrence.

But if consumers misperceive risk or cannot observe whether the quality of service delivered is that contracted for—or if contracting costs are high—consumer willingness to pay for risk reduction is not internalized to the injurer. The tort system may correct this potential market failure by levying a fine over and above the compensatory award. At the extreme, with no internalization through markets and a risk-neutral or fully insured (but perfectly experience rated) defendant, the optimal fine is the value of the injury implied by the victim's willingness to pay for prevention given optimal insurance. The optimal fine is less if (a) markets transmit somewhat, albeit imperfectly, the value consumers place on prevention, or (b) the defendant incurs uninsurable costs of suit, such as loss of time or reputation.[12] In principle, the fine should be paid to the state and refunded as a subsidy to consumers of the risky product. If the fine is paid as compensation to individual victims, it provides more insurance than consumers would buy voluntarily, and hence product prices, which will include the cost of this insurance, would be too high. This distortion of relative prices is avoided if the fine is refunded as a subsidy.

---

[11] See also Daniel A. Graham & Ellen R. Peirce, Contingent Damages for Products Liability, in this issue.

[12] The fine necessary to induce compliance with a negligence standard is less than the optimal fine under strict liability, assuming the due care standard is known and enforced without error, because the defendant's cost function is discontinuous under a negligence rule. See note 71 infra.

Although the current system of compensatory plus punitive damage awards is structurally similar to this ideal two-part system, it is defective in that (1) compensatory awards aim at full compensation for pecuniary plus nonpecuniary loss, regardless of consumers' willingness to pay for such insurance, and (2) punitive awards are judged by defendants' conduct rather than consumers' willingness to pay for prevention, and are paid as compensation to victims rather than refunded to consumers. The full compensation principle implies excessive deterrence and excessive insurance.[13]

The implementation of this theoretically optimal structure of awards requires information on willingness to pay for insurance and prevention that cannot be determined a priori. For this we must turn to empirical evidence.

B.   Evidence on the Demand for Disability Insurance

Theory cannot tell us whether disability lowers or raises the utility of wealth, so cannot tell us whether optimal compensation is less or greater than monetary loss. On this question there are two sources of empirical evidence. Compulsory public coverages—Social Security Disability Income (SSDI), workers' compensation, Medicare and Medicaid—indicate collective choices, given implicit prices which reflect moral hazard and the redistributive features of the funding system. Because participation is compulsory, adverse selection is not an issue. Private first-party coverage indicates what people are prepared to pay for, given the mandatory public programs and tort, at prices that reflect risks of adverse selection and moral hazard in private insurance markets.

1. *Wage Loss.*   Under SSDI, the primary public disability program, income replacement varies with the age, work experience, family size, and predisability earnings of the recipient. Replacement rates for persons eligible for family benefits range from less than 40 percent for those with predisability earnings greater than the taxable maximum ($1,475 per month in 1978) to 86 percent for those earning the minimum wage ($442 per month in 1978). For individuals, comparable replacement levels range

---

[13] Compensation for pain and suffering, if correctly calculated as willingness to pay for prevention, would provide optimal deterrence but excessive compensation. Willingness to pay as a measure of compensation for pain and suffering is discussed in Bryan C. Conley & G. M. Flick, Toward an Objective Valuation of Pain and Suffering (January 1978) (unpublished manuscript, Grad. School Business Administration, Univ. Southern California). They do not distinguish between the optimum award for purposes of compensation and deterrence.

from less than 30 percent to 55 percent. Since these ratios are calculated relative to pretax earnings, replacement of after-tax income is higher.[14]

In 1980, only 21 million people, or 10 percent of the population under sixty-four (19.7 percent of the labor force) had private long-term disability (LTD) income protection. However, 45 percent of civilian workers have private pension coverage which often pays in the event of early disability. Coverage under private LTD and pension plans is less than full income replacement. All private LTD plans limit replacement levels to 60–70 percent of predisability salary and include offset provisions against other coverages, in particular SSDI, to prevent benefits exceeding this limit.[15] For example, for steel workers, whose fringe benefits are more generous than average, replacement income without SSDI benefits averages 0–48 percent of predisability wages, 55–70 percent with SSDI benefits.[16] The apparent gaps in private wage-loss coverage, despite the tax advantages of employment-based plans,[17] appear to reflect the fact that for the majority of workers, SSDI benefits already provide the maximum permitted by private insurers, namely 60–70 percent of predisability, pretax earnings, with offset of SSDI benefits.[18]

2. *Medical Expense.*   Private insurance markets indicate a lack of willingness to pay for unlimited medical care. Forty-eight percent of the population under sixty-five has no private major medical expense protection and private coverage of nursing home or noninstitutional long-term care is negligible.[19] Only 36 percent of major medical plans have unlimited maximum benefits.[20] The prevalence of major medical policies overstates private coverage of long-term medical expense because 76 percent is

---

[14] Charles W. Meyer, Social Security Disability Insurance (1979).

[15] Health Insurance Association of America, New Group Health Insurance Policies (1983).

[16] Lawrence S. Roof, Fringe Benefits: Social Insurance in the Steel Industry 115 (1982).

[17] Insurance premiums and employer contributions to pension funds are a business expense, and interest on such funds is substantially exempt from tax.

[18] Since eligibility for SSDI benefits is not conditioned on other nonearned income, the tendency for a means-tested public insurance program to undermine the demand for private insurance does not apply. Samuel A. Rea, Jr., Private Disability Insurance and Public Welfare Programs, 36 J. Pub. Finance 84 (1981). But a related effect may operate. Private coverage for catastrophic health expense and long-term care is limited (see *infra*). To be eligible for public support for such expenses, the individual must "spend down" to the Medicaid income threshold. Thus the optimum private strategy for long-term institutional care may be to rely on Medicaid, in which case private wage-loss insurance has little value.

[19] Fewer than 4 percent of nursing home expenditures are covered by private insurance. Congressional Budget Office, Long-Term Care for the Elderly and Disabled (1977).

[20] Health Insurance Association of America, Source Book of Health Insurance, 1981–82 (1983).

employment based, so is lost if the disabled person loses his job.[21] But as in the case of wage loss, apparent gaps in private medical coverage may reflect a rational adjustment to public programs. After two years the SSDI recipient is eligible for Medicare, which covers acute care. Long-term institutional care is covered by Medicaid, which pays for over one-third of all long-term-care recipients.

3. *Lump-Sum Payments.*   The only private disability insurance which is not a replacement of specific expenses and thus bears some resemblance to compensation for pain and suffering is accidental death and dismemberment (ADD) insurance, carried by 57 percent of civilian wage earners.[22] These policies typically pay a prespecified sum in the event of a readily identifiable physical injury, such as loss of a limb. But the fact that total ADD contributions by employers and employees are less than 1 percent of total contributions to health benefits is indicative of a relatively low willingness to pay for compensation beyond income replacement and medical expense.

C.   Implications for Compensatory Tort Awards

This evidence from private and collective choices suggests that the tort norm of full coverage of all pecuniary loss plus pain and suffering far exceeds the coverage people are prepared to pay for given the choice. However, private choices may not be adequate evidence, to the extent that they are constrained by consumer misperceptions, moral hazard, adverse selection, and loading charges. If courts have superior access to information, lower costs of controlling adverse selection and moral hazard, or lower loading charges, optimal compensatory tort awards would exceed the revealed choices in private markets. Let us take each of these in turn.

It is often argued that people tend to underestimate very low probability events, which would imply suboptimal insurance purchases. But to purchase appropriate first-party disability insurance, the individual does not need to know his risk from each possible source but only his overall probability of disability, which may be estimated from readily available

---

[21] Because insurance costs of disability act as an impediment to rehiring, all states have second injury funds for workers compensation, but many are very restricted in the scope of prior injuries covered. See Lloyd W. Larson & John F. Burton, Jr., Special Funds in Workers' Compensation (1981) (unpublished manuscript, Cornell Univ., School of Industrial and Labor Relations).

[22] U.S. Department of Commerce, Statistical Abstract of the United States 333 (1981). ADD does not cover disability due to illness, perhaps because illness entails more severe adverse selection and greater moral hazard of fraudulent claims than more readily identified injuries.

data. This overall probability of disability is not negligibly low. In 1978, 10.9 million or 8.6 percent of the noninstitutionalized civilian population aged eighteen to sixty-four were severely disabled (unable to work altogether or unable to work regularly). A further 11.0 million (8.7 percent) reported partial disability (a health condition that restricts the kind or amount of work).[23] Information costs are further reduced by employment-based group programs, which also reduce adverse selection.[24] The fact that employee benefit programs are carefully coordinated with public programs suggests that the apparent gaps in private coverage are due less to ignorance and adverse selection than to a desire by employees and insurers to constrain the moral hazard that would result from higher benefit levels. Moral hazard—which may take the form of reduced effort to prevent injury, exaggerated or fraudulent claims, or increased consumption of health services—obviously limits the coverage private markets can provide efficiently. Thus private contracts limit income replacement to 60–70 percent of predisability earnings, reduce benefits by payments received from public programs to prevent more than full income replacement, impose upper limits on medical coverage, and exclude some services totally.

The tort system seems to have no advantages over private insurance that would justify higher levels of compensation. Juries obviously have inferior information about individual preferences ex ante. As tort coverage is compulsory, it is not subject to adverse selection, but neither are large private employment-based group plans.[25] With respect to moral hazard, the tort system is more exposed than most private coverages. The potential for moral hazard arises because the claimant gains the full benefit of any award, net of his legal fees, whereas the cost is spread among all potential beneficiaries—consumers, in the case of product liability, patients in the case of medical malpractice. To control moral

[23] U.S. Department of Commerce, Statistical Abstract of the United States 336 (1982–83), derived from U.S. Social Security Administration, 1978 Survey of Disability and Work: Databook (1982). A further 1.8 million persons under sixty-five (.9 percent of the under-sixty-five population) were in nursing homes. A 1974 survey showed 12 percent of severely disabled adults received veterans' benefits, so may have been disabled in military service. U.S. Department of Labor, An Interim Report to Congress on Occupational Disease 61 (1980).

[24] For the effect of adverse selection on private insurance markets see Michael Rothschild & Joseph Stiglitz, Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information, 90 Q. J. Econ. 629 (1976). The differential between the price of individual and group health coverage suggests severe adverse selection in the individual market.

[25] This feature of tort might be undermined if contracting out of tort rights were permitted, as suggested in Richard A. Epstein, Medical Malpractice: The Case for Contract, 1976. Am. B. Found. Research J. 87.

hazard, private insurance contracts typically include a coinsurance percentage that is higher the more the policyholder can do to affect the loss. Analogous protections in tort law are the contributory negligence defense (total bar to recovery) and the comparative negligence reduction in the award in proportion to the plaintiff's fault. But contributory negligence has been eliminated in many states in product liability cases, and comparative negligence is of limited scope. Although legal fees constitute a sizable copayment for the plaintiff, their effect is muted by the contingent fee, which is paid only in the event of a positive recovery. In the absence of statutory limits on awards, there is no constraint comparable to the upper limit on coverage found in most private insurance contracts. The private insurance principle of less than full compensation to deter moral hazard is undermined by the collateral source rule, which denies evidence of coverage from other sources, and by disallowing evidence of the taxfree status of tort awards. Finally, with respect to administrative expense, the load on liability insurance policies, which is relevant to determining optimal tort awards, is higher than on first-party insurance.

In conclusion, the tort system has no obvious advantage in terms of information, adverse selection, moral hazard, or administrative cost that would justify setting compensatory awards above the levels of insurance coverage chosen in employment-based group programs.

D. Deterrent Surcharge

The optimal value of the deterrent surcharge, over and above the compensatory award, depends on consumer willingness to pay for injury prevention, given optimal compensation, and on the degree to which this is internalized to injurers either through markets or through uninsured costs of suit.

The best evidence for willingness to pay for injury prevention is labor market data on wage differentials in risky industries, because labor markets reveal insurance and prevention trade-offs made subject to a large measure of cost internalization. Workers' compensation benefits reflect a legislated collective insurance choice in which employers and unions have significant input. Industry-specific collective bargaining provides a mechanism for trading off wage levels, other benefits, and possibly safety measures. Thus, with the exception of SSDI and other public sources of compensation, labor markets reveal choices of injury prevention and insurance when costs are internalized, insurance premiums are widely experience rated,[26] and potential victims, through unions, are reasonably well informed about injury risks.

---

[26] Richard B. Victor, Workers' Compensation and Workplace Safety: the Nature of Employer Financial Incentives (Rand Corp. no. R-2979-ICJ, 1982).

Wage differentials in risky occupations imply a value of life ranging from $500,000 to over $4 million, and a value of $20–$30,000 for nonfatal injuries and illnesses (1980 prices).[27] These labor market estimates obviously cannot be immediately extrapolated to the willingness of potential tort claimants to pay for product or medical injury prevention. First, the theory indicates and the evidence confirms that willingness to pay depends on such factors as age, income, quality of life and risk in the absence of injury, voluntary versus involuntary assumption of risk, and the type of nonfatal injury or illness. Second, private choices are constrained by private budgets. Most public insurance programs incorporate some income redistribution, and it is a normative issue whether the tort system should assess penalties and hence encourage differential prevention, by income level. Third, labor market choices do not internalize all social costs because the public programs, which pay a large share of the costs of disability, are not experience rated. Using a more complete theoretical model to incorporate all social costs, Arthur obtains estimates of the value of saving a life in different contexts that are under $1.4 million (1980 dollars) for all but the youngest age groups.[28]

Even if consumer valuations of tort injuries could be accurately measured, several factors argue against implementing them directly as fines. If the fine is insurable and does not affect the individual firm's premium,[29] then the fine serves no deterrent purpose. If paid to the victim, it provides excessive insurance. If refunded as a subsidy to consumers, the subsidy simply cancels the increment in price due to the cost of insurance, minus administrative costs. If the fine is not insurable, a fine of the magnitude implied by value of life estimates is likely to bankrupt most individual defendants. But as the prospective uninsured loss to the defendant increases, the need for a deterrent surcharge declines, since the optimal fine falls short of the victim's willingness to pay by the amount of any uninsured cost to the defendant.

E. Individual versus Scheduled Awards

The tort system differs from the other major compulsory insurance systems—Social Security Disability (SSDI) and workers' compensa-

[27] W. Kip Viscusi, Risk by Choice: Regulating Health and Safety in the Workplace (1983). The $4 million is estimated at a risk of 6×10^{-4} (6/100,000) per year, and the mean annual risk of nonfatal injury was one in thirty. To see how wage premiums for risk are translated into implicit values for injury, consider the following example. If a worker receives $500 annually in return for a one-in-thirty chance of injury, dividing $500 by $\frac{1}{30}$ yields an implicit value of $15,000 for an injury. This does not mean the worker would accept certain injury for $15,000.

[28] W. B. Arthur, The Economics of Risks to Life, 71 Am. Econ. Rev. 54 (1981). To adjust Arthur's calculations in 1975 dollars to 1980 dollars, I have multiplied by 1.4, the increase in the GNP price deflator.

[29] See note 36 infra.

tion—in determining compensation for each plaintiff individually and after the occurrence of an injury. The only strategy available to defendants (or their insurers) to contain the resulting moral hazard on the part of claimants is legal defense effort. For medical malpractice and product liability, litigation expense by both parties equals the net compensation to plaintiffs.[30] Overall, the defense and plaintiff contribute equally to this total, but the plaintiff spends more, relative to the defense, the larger the stakes.[31]

If tort damage rules set optimal guidelines, but both sides invest in legal effort to maximize their private gain subject to these constraints, will the outcome be socially optimal?[32] Ignoring any precedent and enforcement value of the case, the private benefit to the plaintiff of a higher award is the full social benefit, while the private cost to the defense reflects the social cost ultimately passed on to consumers, through liability insurance premiums and product prices. Thus the outcome should, on average, correspond to the social optimum, provided both parties face similar costs and payoffs to legal effort. We may assume both face a common price of legal effort. The evidence of a declining ratio of defense to plaintiff expenditure then suggests that returns to legal effort decline more rapidly for the defense than for the plaintiff, the more severe the injury. This may reflect a propensity of juries to weigh benefits more heavily than costs when the beneficiaries are identified individuals but the costs are spread, consistent with the familiar tendency to value lives of known individuals more highly than unidentified statistical lives.

This propensity is apparently exacerbated when the defendant is a corporation or an insurance company. For medical malpractice claims, I found an elasticity of awards with respect to the limits of the defendant's insurance coverage of .5–.9 for verdicts, .14 for out-of-court settlements.[33] Peterson finds awards for similar injuries are higher for medical

---

[30] To the extent litigation expense provides information that contributes to the accurate enforcement of tort rules, it should not be viewed entirely as a deadweight cost of moral hazard.

[31] For malpractice claims closed 1975–78, the 64 percent of claims closed for less than $3,000 involved defense expenditure in excess of indemnity paid to plaintiffs. For cases paid over $100,000, defense expense averages less than 10 percent of indemnity. The limited evidence on plaintiff contingent fees shows no comparable decrease with size of stakes and an increase in fee percentage with stage of disposition.

[32] This is a special case of the issue whether common-law rules are efficient. See Paul H. Rubin, Common Law and Statute Law, 11 J. Legal Stud. 205 (1982), and citations therein.

[33] With an elasticity of .5, a $1,000 increase in insurance limits results in a $500 increase in award, holding constant other facts of the case. The estimate for court awards is based on a sample of only thirty-eight claims closed in 1970. The .9 estimate has a t-statistic of 3.19; the .55 estimate, which has a t-statistic of 1.88, includes defense expenditure as an explanatory

malpractice, product liability, and workplace injuries than for automobile injuries, where the defendant is typically an individual and carries lower limits of insurance coverage than most medical or product defendants.[34] The hypothesis that the judicial process tends to sanction moral hazard on the part of severely injured plaintiffs implies that awards for severe injuries will exceed the socially optimal level, consistent with the theoretical and empirical analysis in the previous section.[35]

Even if individualizing awards did not distort optimal compensation on average, it introduces variance that serves no useful purpose. Deterrence almost by definition depends on the *expected* penalty, so involves some element of averaging. If the mean were optimally set, variance would result in overdeterrence of risk-averse defendants. In practice, experience rating of liability insurance premiums typically weights claim frequency more than severity, thereby muting the impact of variance in awards.[36] From the standpoint of compensation, uncertainty reduces the value of an award as insurance to the plaintiff. Equivalently, the mean of a distribution of potential awards provides less utility than would a certain award equal to that mean.

While individualized awards add little to the deterrence and compensation value of the tort system, they do add to the litigation and insurance overhead costs. Like any unappropriated property right, the range of the potential award defines the maximum each party would spend on legal

---

variable. The .14 estimate is based on a sample of 553 claims settled out of court. All three estimates control for number of defendants, severity of injury, age, and income of the plaintiff. Patricia Munch Danzon, The Disposition of Medical Malpractice Claims (R-2622 HCFA, Rand Corp. 1980). Although the insurer can in principle contain this risk by the limits of coverage, some courts have overruled such contractual limits if the insurer declined a pretrial settlement offer.

[34] Mark A. Peterson, Compensation of Injuries: Jury Verdicts in Cook County (Rand Corp. 1983).

[35] The conventional wisdom is that the tort system tends to overcompensate minor injuries and undercompensate severe injuries relative to economic loss. For out-of-court settlements, however, the conventional wisdom is consistent with a rising ratio of plaintiff-to-defense expense, since settlements tend to be lower (higher) the larger the litigation costs of the plaintiff (defense).

[36] This reflects in part the belief of liability insurers that the size of award, at some threshold, is beyond the control of the defendant, in part the purely statistical fact that since large awards are relatively infrequent, they have little statistical credibility. Reflecting this, both medical malpractice and product liability use a finer rating system for basic losses (claims below some threshold, $25,000 in 1980) than for excess losses. For example, basic Insurance Services Office (ISO) rates for medical malpractice reflect eight medical speciality classes and over fifty territories, but only two excess limits factors. To the extent merit rating of medical malpractice policies exists, it is independent of size of award. Patricia M. Danzon, Liability Insurance and the Tort System: The Case of Medical Malpractice (Working Paper No. E-83-14, Hoover Inst. 1983). Small losses also receive more weight in workers' compensation experience rating. Victor, *supra* note 26.

effort to influence the outcome. Further, individualization adds to the variance and positive skewness of the distribution of awards and hence to the variance of insurance losses. In principle, this source of risk can be eliminated by pooling. In practice the number of policyholders in many states is too small to eliminate such risk.[37]

F.   The Collateral Source Rule

The collateral source rule prohibits evidence to the jury of the plaintiff's compensation from other private and public programs. Elimination of the collateral source rule, to permit reduction of tort awards by the amount of collateral benefits, has been enacted in several states for medical malpractice cases and proposed for product liability.

The main argument in favor of collateral source offset is to prevent double compensation: "The idea of a windfall runs counter to the basic end of tort law, which is to make the plaintiff whole, not to overcompensate him. . . . the aim should be to assure the plaintiff fair compensation from available sources, but no more."[38]

But Posner has defended the traditional no-offset rule as necessary for optimal deterrence.[39] Further, he argues that any double recovery is no windfall to the plaintiff since he paid for his first-party coverage and could, presumably, have opted for a cheaper policy that excluded coverage of tort claims or assigned subrogation rights in such claims to his first-party insurer. One may add, at least in contexts where the victim and defendant are in a buyer/seller or employee/employer relation, that the victim has also indirectly paid for the compulsory insurance provided through the tort system, so is entitled to what he paid for.

In a world with perfect information and costless contracting, the Coase theorem implies that the collateral source rule is neither necessary nor sufficient for optimal deterrence or optimal insurance. Consider the case of an injury that entails a purely monetary loss, fully covered by first-party insurance, and for which the tort rule of full compensation provides optimal insurance and deterrence. If the traditional collateral source rule of no offset applies, the tort award provides optimal deterrence, and a subrogation action by the first-party insurer against the insured (the plain-

---

[37] For example, year-to-year changes in the mean medical malpractice award exceed 100 percent for half the states in the 1975–78 period and range from −90 percent to +1,000 percent. Thirty-five states had fewer than one hundred paid claims per year. Danzon, *supra* note 6.

[38] American Bar Association, Report to the Commission on Medical Professional Liability 147 (1977).

[39] Richard A. Posner, Economic Analysis of Law (2d ed. 1977).

tiff in the tort suit) can eliminate double compensation.[40] On the other hand, if the tort rule is full reduction of the award by the amount of collateral compensation, optimal compensation is achieved through the first-party insurance, and a subrogation action by the first-party insurer against the tortfeasor provides optimal deterrence. Thus, in either case, two actions are necessary but together are sufficient for optimal insurance and prevention. Alternatively, the subrogated insurer may simply receive part of the proceeds from any judgment or settlement.

In practice, under the traditional rule of no offset, subrogation mitigates but does not fully eliminate double recovery because contracting out is not costless. The law recognizes subrogation rights of medical insurers against tortfeasors or their liability insurers,[41] and the prevalence of sub-rogation clauses in medical policies indicates public preference to elimi-nate double coverage. But life and accident insurance has generally been denied subrogation rights, on the grounds that it provides "personal" insurance rather than "indemnity" insurance; that is, life and accident policies allegedly do not compensate for explicit economic losses. As Kimball and Davis point out, there are significant indemnity aspects in all forms of life and disability insurance. They conclude: "[W]here there is an insurance against a loss measurable in economic terms, the insurer should be legally subrogated quite as readily as in fire or collision insur-ance, once there has been full indemnification of the insured. Denial of legal subrogation should reflect the lack of adequate indemnification to the insured in the individual case, and nothing more.[42]

But to allow subrogation only after full indemnification of the insured simply invites litigation over ill-defined rights, since indemnification for personal injury can never be complete. Unrestricted freedom to contract for subrogation in life and accident policies should be allowed, both to discourage the moral hazard created by double coverage and to permit individuals to choose their preferred level of compensation.

Subrogation rights found in public programs are mixed. The Federal Medical Care Recovery Act grants the federal government a right to

---

[40] Optimal compensation might be achieved without the supplementary subrogation ac-tion if the first-party insurance simply excluded coverage of any injury with recovery against a third party. But such contracts tend to provide inferior insurance because of the uncer-tainty of tort recoveries and are less common than contracts providing subrogation rights.

[41] Spencer L. Kimball & Don A. Davis, The Extension of Insurance Subrogation, 60 Mich. L. Rev. 841 (1962). Most first-party automobile medical coverages (which pay the policyholder's medical costs as an endorsement to his liability insurance policy) provide for subrogation against third parties.

[42] *Id.* at 859–60.

recover from a tortfeasor for medical expenses it incurs.[43] Workers' compensation employers or their insurers have been explicitly accorded subrogation rights by statute in most states. However, SSDI does not have subrogation rights.

Although the collateral source rule is irrelevant when subrogation is permitted and contracting costs are zero, the optimal rule in practice depends on the costs of contracting out of double compensation and of enforcing subrogation. One obvious guiding principle is that mandatory coverage should pay in full and be denied subrogation, whereas voluntary first-party coverages should be free to include enforceable subrogation clauses, thereby enabling individuals to contract out of double coverage or supplement the mandatory coverage if they desire. This principle argues for retaining the traditional collateral source rule, since tort awards are a form of compulsory insurance, but granting all first party coverages subrogation rights.

This rule allowing subrogation to voluntary coverages provides no guidance for coordinating tort with other mandatory coverages, in particular, SSDI. If tort awards were paid in periodic installments rather than a lump sum, then any future SSDI benefits could be reduced by the amount of the tort compensation for earnings loss, eliminating excess compensation while preserving deterrence.[44] Providing for SSDI offset when tort awards are paid in a lump sum is less easy. Granting SSDI subrogation against the plaintiff at a later date may be unenforceable. If the SSDI offset must be made at the time of the tort judgment, the simplest solution is to reduce the tort award by the amount of the expected SSDI benefits, but this undermines deterrence. The alternative of granting SSDI subrogation rights against the tortfeasor yields superior deterrence and equivalent compensation, but with higher litigation expense. Such expense may be significant in the great majority of cases that settle out of court, typically for somewhat less than the potential verdict. The subrogation action must then determine whether the subrogated insurer should be entitled to full recovery or only to the proportion represented by the settlement relative to the potential verdict.[45] Coordination of legal effort by the plaintiff and

[43] Rex Capwell & Thomas E. Greenwald, Legal and Practical Problems Arising from Subrogation Clauses in Health and Accident Policies, 54 Marq. L. Rev. 255, 281 (1971).

[44] Social Security Disability Income has similar offset provisions against state workers' compensation programs. Periodic payment of an amount fixed at the time of the judgment avoids the moral hazard of periodic payment of contingent expense discussed in Samuel A. Rea, Jr., Lump-Sum versus Periodic Damage Awards, 10 J. Legal Stud. 131 (1981). Optimal insurance would provide for disability payments for the duration of the victim's life, with some lesser payment to survivors in the event of early death, as in the California Medical Injury Compensation Reform Act, Code Civ. Proc., § 667.7.

[45] The majority of courts use a prorata division, but some assign full compensation to the insured and some full compensation to the insurer. Robert E. Keaton, Insurance Law (1971).

subrogated insurer may also add costs. Thus efficient coordination of SSDI and tort benefits is an additional reason for periodic payment of tort awards.

G.   Conclusion

This analysis indicates that the optimal structure of tort awards would be a schedule providing[46]

compensation for medical expense such as that found in group major medical policies;

compensation for wage loss up to 70 percent of predisability, pretax earnings (full replacement of after-tax earnings);

no offset of private insurance coverage but unrestricted subrogation rights of private insurers against tortfeasors and their insurers;

reduction of SSDI benefits by the amount of tort recovery for wage loss;

a schedule of compensation for pain and suffering for serious injuries only; and,

consistent with the principle of eliminating uncertainty, statutory standards for determining inflation, interest rates, and wage growth parameters.

In addition, an uninsurable fine may be appropriate in cases of severe injury, when (1) consumer underestimate of risk, fraud, or breach of explicit or implicit contract are at issue, or (2) defendants are fully insured. Where consumers are adequately informed or the defendant has incurred significant loss of time or reputation in defending the suit, deterrence incentives may be adequate without a fine. The fine should be paid to the state and used to defray the public costs of the judicial system, thereby internalizing some of these costs to activities that cause injuries. Although this reduces the plaintiff's incentives to litigate, the effect may be negligible since a fine only applies in cases of severe injury where the compensatory award is large.

Note that, to the extent the purpose of punitive damage awards is deterrence, this analysis of the optimal deterrent surcharge or fine applies directly: the optimal punitive award is simply the optimal fine.[47]

---

[46] This is not intended to preclude individual contracting for amounts different from the basic schedule, as advocated in Epstein, *supra* note 25.

[47] Obviously, if compensatory awards continue to compensate for pain and suffering, no fine is necessary. For an analysis of punitive damage awards, see Symposium on Punitive Damage Awards, Discussion, 56 S. Cal. L. Rev. 155 (1982).

II.   DURATION OF LIABILITY

A.   Risk Associated with Latent Injuries

A common distinguishing feature of medical malpractice and product liability is the long duration of potential liability because of latent injuries and long-lived products. Regardless of the assignment of liability, delay between the triggering event and the manifestation of injury creates uncertainty in preventing and compensating for injuries. First, delay creates technological uncertainty in establishing the connection between the triggering event and the injury. Carcinogenic chemicals and drugs with long latent side effects are obvious examples. Second, delay introduces a financial risk in prefunding future compensation.[48]

Tort law has increasingly extended liability to protect victims of latent injuries. For medical malpractice, many courts have adopted liberal discovery rules, which toll the running of the statute of limitations until the injury has been, or with reasonable diligence should have been, discovered. For product liability actions in tort, the statute does not begin to run until the occurrence or manifestation of the injury, which may be many years after the manufacture and initial sale of the product or exposure to the toxic substance. Since 1975, thirty-eight states have shortened their medical malpractice statutes of limitations or set an outer limit on the period allowed for discovery, but often with exceptions for minors and latent injuries. McGovern identifies ninety-eight statutes in forty-eight states that can be considered product liability statutes of repose.[49]

In general, court rulings on the constitutionality of such statutes[50] have not appreciated their effects on the efficiency of the tort system, some of which are addressed here. I ignore the fact that delay leads to decay of evidence, blurs the chain of causation, adds multiple defendants, and hence increases litigation expense. I focus instead on the effects of changing and uncertain liability rules on deterrence, risk allocation, and regulatory costs.

First let us assume that standards are changing but known with certainty. For example, there may be an upward trend in compensable damages over time or a technological advance that reduces the cost of preventing injuries. The effect of such changes is that the optimal level of care becomes ambiguous. Prevention costs occur at time 1 but damages

---

[48] The financial risk arises because of uncertain timing, rather than delay per se. A certain time schedule of future payments can be matched with the maturity schedule of financial assets, thereby eliminating financial risk.

[49] McGovern, *supra* note 5.

[50] Note 3 *supra*.

occur at time 2. Under a negligence rule, an action that is not negligent by the standards of time 1 is negligent if costs or damages are valued at time 2. Similarly, under a strict liability rule the optimal level of care becomes ambiguous.

If future liability for latent injuries is fully internalized into the prevention and insurance decisions at time 1, consumers at time 1 pay for the higher safety and insurance standards of consumers at time 2. Provided the same individuals are involved throughout, this is no cause for concern. But with long latent injuries there are overlapping generations. The consumers who purchase the product or incur the exposure at time 1 are not the same as those who set liability rules at time 2, although obviously there is some overlap. Then with full internalization, long statutes force a transfer from the early to later generation.

Alternatively, producers may choose not to prefund latent liability at time 1 but rather to face liability costs when they occur at time 2. But in a competitive industry with free entry these costs cannot be passed on to consumers at time 2. Established firms that try to pass on current costs arising out of past liability will be undercut by new entrants. So the cost of latent injuries will fall initially on producers and then on victims, in the event producers go bankrupt.

The intergenerational transfer is a more serious issue when the injury itself is latent than it is in the case of a long-lived machine that may cause injuries over an extended period, but where the injury itself is immediately manifest. With the latent injury, the only transaction is between producers and consumers at time 1. If standards change, there is no way later generations can be made to pay for their tastes. With the long-lived machine, on the other hand, the rental or resale price in each period will reflect the expected costs of injuries in the period. If liability costs change due to rising standards of compensable damages, this can be reflected in prices to later generations of consumers. But if the increase in liability costs results from technological advance and a retroactively applied shift in the due-care standard, then in competitive markets the older, less safe machine will not be able to recoup its higher liability costs.[51] This forced obsolescence is no different from obsolescence due to any change in technology or tastes, except that retroactively applied liability rules curtail a potentially longer economic life for those consumers willing to take

---

[51] In Beshada v. Johns-Manville Products Corp., 90 N.J. 191, 447 A2d 539 (1982), the New Jersey supreme court ruled that a manufacturer can be held liable for its failure to warn even if the hazard in question was unknowable at the time of the product's sale. See Gary T. Schwartz, New Product, Old Product, Evolving Law, Retroactive Law, 58 N.Y.U.L. Rev. 796 (1983).

the higher risk in return for a lower price. But this is simply a special case of the general point that liability rules mandate uniform standards, thereby tending to deprive consumers of the freedom to choose riskier products or lower compensation. Thus one argument for a state-of-the-art defense for long-lived products is that it permits consumers more options. But if consumers are homogeneous in their preference for safety, courts that apply current standards of care retroactively are merely enforcing what informed markets would do otherwise.

Now consider the more realistic case, where liability rules—the scope of liability, standard of care, rules of evidence, and size of damage awards—not only change but are unpredictable. Uncertainty about legal rules creates a risk additional to the unavoidable technological and financial risks intrinsic to latent injuries. This has two important implications for liability insurance (see Appendix B). First, sociolegal risk creates parameter uncertainty by destroying the insurer's ability to predict the loss distribution with any accuracy. Experience of the most recent policy years is not fully "mature," but the more mature experience of older policy years rapidly becomes obsolete in a nonstationary environment.[52] But even if statistical analysis could estimate past trends accurately, there is no guarantee the future will replicate the past.

Second, a long statute of repose induces positive correlation among policyholders. The distribution of the expected loss per policyholder depends not only on the probability of an injury, but also (1) on the conditional probabilities, given an injury, that a claim is filed; (2) that an award is made; and (3) on the size of the award. These last three factors depend on the sociolegal climate common to all insureds. The longer the statute of repose, the greater the weight of sociolegal risk relative to individual risk, the greater the covariance among policyholders, and hence the greater the variance in the mean outcome for the insurance portfolio. At the limit, perfect correlation among outcomes for all policyholders is equivalent to insuring one single policyholder. Long statutes of repose in a volatile legal environment therefore reduce the financial gains from pooling, relative to self-insurance.

---

[52] The longer the statute of repose, the greater the backlog of claims that will be filed in response to a pro-plaintiff legal change, and the greater the amplitude of transitory cycles in frequency of claims filed per year. If the statute of repose is ten years, it requires ten years to determine whether an increase in the number of claims filed in the first year represents accelerated reporting of a given number of claims—in which case estimates of unreported claims should be decreased—or an increase in underlying frequency—in which case estimates of unreported claims should be increased.

B.   Market Responses to Residual Risk

First, a digression is necessary to describe the types of liability insurance contract, which determine the allocation of risk between the insurer and the policyholder. Traditionally, medical malpractice and product liability insurance have been written on an occurrence basis. The typical medical malpractice occurrence policy written in policy year 1984 covers claims arising out of practice in 1984, no matter how far in the future the claims may be filed.[53] A "claims made" policy covers claims *filed* in the policy year, arising out of prior practice; claims filed in subsequent years can be covered by a reporting endorsement, the availability but not the price of which is guaranteed when the basic policy is purchased. Thus, to have equivalent insurance with claims made, the policyholder must reserve or self-insure for the reporting endorsement at the time he purchases the initial policy. The typical product occurrence policy covers injuries occurring in the policy year. Neither the time of manufacture of the product nor the time at which the claim is made is relevant. As in the case of claims made, since the insurer has more experience on which to base the premium, the policyholder is exposed to more pricing risk under this quasi-occurrence product policy than under the medical malpractice occurrence policy.[54]

The medical malpractice occurrence coverage forces full prefunding, whereas the product occurrence policy and the claims made permit "pay as you go." To predict the appropriate premium for medical occurrence coverage, the insurer must predict claim frequency and severity for the duration of the discovery period or statute of repose, plus any lag in disposition beyond that, up to fifteen years. For claims made, the insurer must predict liability retrospectively for the discovery period. For the occurrence product policy, in the absence of a statute of repose, retrospective prediction is indefinite and prospective prediction is for the statute of limitations, which does not begin to run until the discovery of the injury.

Residual risk for a given line of insurance, due to small risk pool, parameter uncertainty, or positive correlation among policyholders, is not necessarily fatal to diversification through market insurance. Di-

---

[53] Since a policy covers one year, policies written in 1984 cover incidents occurring January 1984–December 1985. Thus a policy year spans two calendar years.

[54] This applies particularly where all units of a product line are subject to similar risks. Some recent court rulings have interpreted product liability policies, like medical malpractice occurrence policies, to run from the injury-causing act of the insured rather than from the manifestation of the injury.

versification may occur within a multiline firm, writing different lines of insurance with negatively correlated experience, or through stock markets. The capital asset pricing model (CAPM) implies that residual insurer risk is nondiversifiable *only* if it is systematic, that is, underwriting losses are highly correlated with losses for the market as a whole (high betas). But if underwriting losses are uncorrelated with the market, insurance premiums should reflect expected claim costs with little markup for risk.[55]

There is some evidence that, contrary to these predictions of CAPM, the total variance of a firm's returns, including the nonsystematic component, is a significant determinant of the price of capital.[56] In that case, or if betas are high, the competitive insurance premium would include a markup above expected costs. If the markup must be sufficient to reduce to a predetermined level the probability that claim costs exceed premiums, then the markup is proportional to the residual standard deviation of the mean, σ_c, which increases with the statute of limitations. The factor of proportionality increases with the desired safety margin. For example, if the desired probability of premium inadequacy should not exceed .01, the markup is $2.33\sigma_c$ for large risk pools, more for smaller groups. Evidence from medical malpractice rate hearings suggests that a standard deviation at least equal to the mean would be a conservative estimate. That would imply a "very safe" premium rate three times the expected claim costs, plus overhead. I do not wish to imply that risk premiums of this magnitude are built into malpractice or product liability rates.[57] I do contend that, with a long duration of liability, a substantial markup for risk may be warranted, even in competitive insurance markets.

If premiums include a markup for nondiversifiable risk, the expectation is that ex post there will be net transfer from policyholders to equity owners. This raises the price of market insurance relative to self-

[55] The CAPM implies a competitive insurance premium, $P = [E(c) - \lambda\sigma_{m,c}]/(1 + R_f)$, where $E(c)$ is expected costs, λ is the market risk premium, $\sigma_{m,c}$ is the covariance between the market return and claims, and R_f is the risk-free rate of return. Underwriting betas, by line, cannot be estimated by standard methods, even for stock insurers, because the typical firm's overall beta compounds returns on multiple lines of insurance, investment and noninsurance operations. Raymond D. Hill, Profit Regulation in Property-Liability Insurance, 10 Bell J. Econ. 172 (1979); William B. Fairley, Investment Income and Profit Margins in Property-Liability Insurance: Theory and Empirical Results, 10 Bell J. Econ. 192 (1979); Patricia Munch & Dennis E. Smallwood, Theory of Solvency Regulation in the Property and Casualty Insurance Industry, in Studies in Public Regulation (Gary Fromm ed. 1981).

[56] Haim Levy, The CAPM and Beta in an Imperfect Market, 6 J. Portfolio Management 5 (1980).

[57] The formula used by the ISO to calculate advisory rates incorporates only a 5 percent markup over losses plus expenses, for profit and other contingencies. But other components of the formula—in particular, projected losses—are sufficiently judgmental to accommodate an implicit allowance for risk.

insurance and increases incentives for risk retention. There is consider-
able evidence to support this hypothesis. In medical malpractice, in re-
sponse to the large premium increases in 1975, physician-owned mutuals
were formed and have expanded to cover over 40 percent of the market in
1982. A substantial fraction of hospital insurance is written through cap-
tive companies. The leading commercial carrier and several smaller car-
riers have replaced the traditional occurrence policy with a claims-made
policy, which transfers from the insurer to the insured the risk associated
with reserving for late claims.[58]

For product liability, measuring the market shares of self-insurance and
commercial insurance is impossible, because product liability is typically
written as part of a comprehensive general liability policy and, more
fundamentally, because the market itself—those potentially exposed to
product liability—is ill defined. But risk retention is obviously significant.
Trade association surveys, admittedly nonrandom samples, show up to 30
percent of firms without insurance, increased deductibles over time, and
lower limits of coverage relative to expected losses.[59] Large firms often
purchase retrospectively rated policies and many have formed captive
insurers or self-insure, despite the tax advantages of commercial insur-
ance.[60] Smaller firms, for whom the risks and fixed capital costs of self-
insurance or captive insurance are relatively high, have obtained special
legislation to form risk pools.[61]

C. Implications of Long Statutes of Repose for the Tort System

The statute of repose defines the duration of liability shifting. The test
for an efficient statute involves weighing, at the margin, any efficiency
gains and losses in risk allocation, deterrence, and overhead costs. Obvi-

---

[58] Since insurers presumably face lower costs of estimating and reserving for future
liability than do individual policyholders, one attraction of claims made apparently is its pay-
as-you-go feature: there is no requirement to predict and prefund future liability. This is less
costly to policyholders in industries with significant barriers to entry and where insurance
premiums are not individually rated, which may explain why claims made has been adopted
in medical malpractice but not product liability.

[59] U.S. Department of Commerce, Interagency Task Force on Product Liability, Final
Report (1978). Over 30 percent of firms in the machine tool industry either were not insured
or carried a deductible averaging $80,000. Problems Associated with Product Liability:
Hearings Before the Subcomm. on Consumer Protection and Finance of the House Comm.
on Interstate and Foreign Commerce, 96th Cong., 1st Sess., June 26, September 27, and
October 16, 1979, at 35.

[60] Self-insurance reserves are not tax deductible until paid out as claims. Premiums paid
to a captive insurer are tax deductible only if there is a transfer of risk, the operational
definition of which is currently under litigation. Business Insurance, August 1, 1983.

[61] Product Liability Risk Retention Act of 1981, 15 U.S.C. § 3901.

ously there is no unique optimal statute for all contexts. My purpose here is to point out how the gains from attempting full internalization through a long statute fall as time elapses from the triggering event.

Consider first risk spreading, which has been one of the primary justifications for extending the liability of producers.[62] The argument tends to ignore risk spreading through first-party insurance; it presupposes that the producer is fully insured and can pass on the cost of this insurance in the product price, such that consumers pay for their own insurance. In this ideal situation there is no burden on producers and no transfer between generations of consumers. The evidence above suggests that such costless, distribution-neutral insurance is not feasible for latent injuries subject to sociolegal risk. If future injury costs are fully prefunded, either through an occurrence policy written at the time of the triggering event, or through a quasi-occurrence or claims-made policy where the producer prefunds the expected cost of a reporting endorsement, the early generation of consumers pays for the expected standards of liability and damages adopted by later generations and for uncertainty as to those standards. If this future liability is not prefunded, the intergenerational transfer is avoided. But to the extent current insurance costs to cover prior liability cannot be passed on, these costs fall on producers or injured victims. Thus the allocation of risk achieved by long statutes of repose is not necessarily preferable to that achieved with a short statute, given the availability of first-party private and public insurance.

With respect to deterrence, if the time path of liability rules entails an anticipated expansion of compensable damages or retroactive application of new knowledge, then the legal changes force a transfer from early to later generations. Such a transfer cannot be judged inefficient but may be considered undesirable, if wealth is rising over time. Where liability is not only expanding but also uncertain, such that insurance premiums include a markup for nondiversifiable risk, this markup is equivalent to a tax on activities with potential latent injuries and long-lived capital goods, if producers insure or self-insure. But if the response to high insurance costs is to go bare, such that the expected cost of injuries is limited to assets that can be attached in bankruptcy proceedings, the prevention incentives that remain may be suboptimal. The incentive to go bare and risk or even plan bankruptcy is greater the less the specific or brand-name capital at

---

[62] For example, Justice Traynor's concurring decision in Esola v. Coca Cola Bottling Co., 24 Cal. 2d 453, 150 P.2d 436 (1944), "The cost of an injury . . . may be an overwhelming misfortune to the person injured, and a needless one, for the risk of injury can be insured by the manufacturer and distributed among the public as a cost of doing business." See also Greenman v. Yuba Power Products, Inc., 59 Cal. 2d 57, 27 Cal. Rptr. 697, 377 P.2d 897 (1963).

stake.[63] Consistent with this prediction, small firms, which arguably have less brand-name capital than large firms, are more likely to go without product liability insurance,[64] and very few physicians go bare. In general, if there is sufficient uncertainty about liability rules to generate significant nondiversifiable insurance risk, the value of such liability as a guide to prevention is questionable.[65]

In addition to the direct impact on prevention and risk allocation, the uncertainty created by long statutes of repose has generated other indirect costs. In malpractice, disagreement over the appropriate price of insurance has led to regulation of insurance premiums to levels deemed inadequate by commercial companies and to their total withdrawal from several states.[66] By contrast, product liability, which so far is not subject to rate regulation, experienced premium increases at least as great as those proposed for medical malpractice but no comparable lack of availability. The Interagency Task Force on Product Liability cited judgmental insurance rate making as a major cause of the product liability problem. Its proposed solutions include more comprehensive data collection and regulatory monitoring of rates to ensure that they are "fair, nondiscriminatory and reasonably related to product risk." The analysis here suggests that even if the full universe of potential data is available, rate making will optimally involve an element of judgment commensurate to the relative importance of sociolegal risk.[67] Where such risk is large, uncertainty about fair rates cannot be resolved statistically. Regulation of rates to the lower end of the feasible range is likely to induce carrier withdrawal and lack of availability as occurred in medical malpractice.[68]

[63] Munch & Smallwood, *supra* note 55.

[64] The greater propensity of small firms to go bare may also reflect a relatively higher price of insurance.

[65] See James A. Henderson, Jr., Coping with the Time Dimension in Products Liability, 69 Cal. L. Rev. 919 (1981). The bankruptcy filings of three major corporations and implied inadequacy of funds, through either insurance or self-insurance sources, to pay asbestos claims is apparent evidence that in this context imposing liability does not deter, either because the injuries were unforeseeable, the liability was unforeseeable, or the bankruptcy option was cheaper. The net result is an excessively costly form of insurance.

[66] Patricia Danzon, The Medical Malpractice Insurance Crisis Revisited: Causes and Solutions (Working Paper E-83-11, Hoover Inst., July 1983).

[67] With pooled information and rate making in concert, errors in parameter forecasts will be correlated and insurer insolvencies will tend to occur together. Emilio Venezian, Insurer Capital Needs under Parameter Uncertainty, 50 J. Risk & Insurance 19 (1983).

[68] To ensure availability, most states established a joint underwriting authority (JUA) which requires commercial insurers to write medical malpractice as a condition of writing other lines of insurance in the state. Typically, JUA losses may be recouped by a tax write-off or a surcharge on premiums in other lines of insurance. While some subsidy may in principle be justified to offset the risk tax implicit in insurance premiums, this is an inefficient method of effecting such a subsidy.

D.   Conclusion

I have argued that unlimited producer liability for latent injuries and long-lived products tends to create costs of inefficient risk allocation, overdeterrence, and regulatory conflict that may outweigh any benefits from attempting full cost internalization. If producer liability is to be curtailed, one option is simply to apply a short statute of repose, running from the occurrence of the injurious act (or date of first sale in the case of capital goods) rather than from the manifestation of injury. The statute should be shorter the more volatile are sociolegal rules. Injuries discovered after the running of the statute would be borne by the victim's first-party insurance or public programs. Since these are pay-as-you-go programs covering all injuries regardless of cause, the issue of nondiversifiable risk does not arise. The presumption of some judicial rulings that manufacturers are better able than consumers to spread risk is at best of doubtful validity, but almost certainly false when sociolegal risk is at issue. Let us allow that risk misperception and adverse selection lead to suboptimal private insurance for those not eligible for employment-based group programs. That this problem applies to all injuries, not just to the small fraction potentially eligible for tort compensation, argues for a general public program such as SSDI rather than for extending the compulsory insurance through tort to cover at most a small fraction of catastrophic injuries. It is ironic and unfortunate that current rules of eligibility for SSDI are based on recent employment and exclude precisely those for whom private insurance markets function relatively poorly.

An alternative solution to the latent injury problem is to shift liability to the government. In some states, compensation funds already pay malpractice awards beyond a statutory threshold—typically $100,000 per defendant. Proposals for special government funds for occupational disease are proliferating; but, again, they perversely single out the employed who, among all victims of toxic exposure, are most likely to have alternative private or workers' compensation coverage. The main advantage of government intervention is to prevent the forced prefunding of future liability, the resulting transfer from early to later generations, and the risk tax on the early generation (ignoring planned bankruptcy). But this purpose is served by simply curtailing liability through a statute of repose.

The disadvantages of government funds are obvious. With respect to deterrence, government funds tend to protect producers from efficient as well as inefficient impositions of liability. With respect to compensation, theory and evidence from other government programs suggest that when benefit levels and eligibility criteria are set through the political process for a specific, identifiable group of beneficiaries, the moral hazard—or

tendency to award excessive levels of insurance—may be greater than through the tort system.

A full analysis of the occupational disease problem is beyond the scope of this paper. The analysis here provides an additional reason to extinguish tort remedies against product suppliers or employers. To encompass diseases, such as cancer, that have at least a twenty-year latency period would require a statute of repose at least that long. The risk costs of such a statute probably outweigh any benefits, given the volatility of tort standards. This logic applies to a lesser degree to workers' compensation liability, because compensation benefits and eligibility criteria are set by statute and can therefore be more readily constrained. Of course the problems of prolonged liability are avoided if liability is imposed on the employer at the time the injury is manifest. But this eliminates all deterrence and exposes victims to risk. Because prior liability costs cannot be passed on in current product prices, the person with prior toxic exposure may face difficulties in obtaining new employment, unless costs are shifted to second injury funds. A superior solution for deterrence and risk allocation may be to prorate liability over all employers contributing to the exposure; to limit benefits to levels prevailing at the time of exposure, adjusted upward by the rate of return on reserves between exposure and manifestation; and to rely on second injury funds or first-party insurance and SSDI in the event none of the responsible employers are available.

APPENDIX A

OPTIMAL DAMAGE AWARDS

The model applies in any market context where consumers may be injured by product failure that depends only on the care taken by producers. Assume that expenditure on prevention affects the probability but not the size of loss and that each consumer buys just one unit of the product.[69] Insurance for first-party or liability losses is available, with perfect experience rating and a proportionate loading charge. The following notation is used:

$$V(B) = \text{consumer's utility of initial wealth, } V' > 0, V'' < 0;$$
$$p(r) = \text{probability of injury, } p' < 0, p'' > 0;$$
$$r = \text{quality (prevention) per unit;}$$
$$s = \text{product price;}$$
$$c(r) = \text{production cost per unit, } c' > 0, c'' < 0;$$
$$L = \text{monetary loss to consumer if injury occurs;}$$
$$M = \text{first-party insurance coverage bought by consumer;}$$

---

[69] This abstracts from the effects of liability rules on the level of activity. See Steven Shavell, Strict Liability versus Negligence, 9 J. Legal Stud. 1 (1980).

γp = premium rate per dollar of first party coverage, where $\gamma \geq 1$ is the loading charge;

$h(p)$ = consumer's perception of p, $h' > 0$, $h'' < 0$;

$U(A)$ = producer's utility of initial wealth, $U' > 0$, $U'' < 0$;

D = damages paid by producer if found liable;

Q = liability insurance coverage bought by producer;

λp = premium rate per dollar of liability coverage, where $\lambda \geq 1$ is the loading charge;

μ = Lagrange multiplier.

Subscript 0 denotes the state in which an injury occurs. Subscript 1 denotes the state in which no injury occurs. Subscripts f, s, and n denote first-party, strict, and negligence liability, respectively. Initially, consumers are assumed to be fully informed.

FIRST-PARTY LIABILITY

If consumers are fully informed, competitive markets induce producers to choose the level of safety (r) and product price (s) to maximize expected utility of consumers, $E(V)$, subject to maintaining an opportunity level of utility, U^c, determined by the producer's alternative use of time.[70] Consumers select first-party insurance coverage (M), given the price per dollar of coverage, p. Informed markets thus solve the following optimization problem:

$$\max_{M,s,r} \phi = (1 - p)V_1[B - s - \gamma pM] + p \, V_0[B - s - \gamma pM - L + M]$$
$$+ \mu \{U[A + s - c(r)] - U^c\}. \tag{A1}$$

Maximization with respect to M, s, and r yields

$$\gamma \, \bar{V}' = V_0', \tag{A2}$$

$$\bar{V}' = \mu \, U', \tag{A3}$$

$$c' = -p' \left(\frac{V_1 - V_0}{\bar{V}'} + \gamma M \right), \tag{A4}$$

where $\bar{V}' = (1 - p) \, V_1' + pV_0'$.

Rewriting equation (2),

$$\frac{V_1'}{V_0'} = \frac{1 - \gamma p}{\gamma - \gamma p} \leq 1 \quad \text{as } \gamma \geq 1, \tag{A2'}$$

shows that optimal coverage does not fully equate the marginal utility of income if the insurance premium contains a proportionate load ($\gamma > 1$). Equation (A4) shows that if injury reduces the utility of income ($V_1 > V_0$), optimal prevention (r_f^*) may exceed the optimal level with risk neutrality ($c' = -p'L$), even with full insurance coverage of any monetary loss.

---

[70] If the producer has monopoly power, U^c includes some rent, but the structure of the problem is not affected.

Strict Third-Party Liability

Under a rule of strict third-party liability, an omniscient benevolent dictator would choose the damage award (D), defendant's liability insurance coverage (Q), prevention (r), and product price (s) to maximize the consumer's expected utility, subject to maintaining the producer's opportunity level of utility, U^c:

$$\max_{D,Q,s,r} \phi = (1 - p)V_1[B - s] + pV_0[B - s - L + D]$$
$$+ \mu \{(1 - p)U_1[A + S - c(r) - \lambda pQ] \tag{A5}$$
$$+ pU_0[A + s - c(r) - \lambda pQ - (D - Q)] - U^c\}.$$

Maximization with respect to D, Q, s, and r yields:

$$V_0' = U_0'\mu \tag{A6}$$

$$\lambda \bar{U}' = U_0' \tag{A7}$$

$$\bar{V}' = \mu \bar{U}' \tag{A8}$$

$$c' = -p'\left(\frac{V_1 - V_0}{\bar{V}'} + \frac{U_1 - U_0}{\bar{U}'} + \lambda Q\right), \tag{A9}$$

where $\bar{U}' = (1 - p)U_1' + pU_0'$.

Equations (A6), (A7), and (A8) together imply:

$$\frac{V_1'}{V_0'} = \frac{U_1'}{U_0'} = \frac{1 - \lambda p}{\lambda - \lambda p} \leq 1 \text{ as } \geq 1.$$

The optimal damage award, D^*, provides the level of insurance consumers would choose to buy given the load of the producer's liability insurance. Thus $D^* \gtreqqless M$, as $\lambda \lesseqqgtr \gamma$.

Negligence

Under a negligence rule, the producer is liable only if he fails to meet the due-care standard, r_n^*. The social welfare function is given by equation (A1) for $r < r_n^*$ and by equation (A5) for $r > r_n^*$. If U is state dependent, the producer cannot fully insure against the loss. The social welfare function is discontinuous, with a vertical jump at r_n^* because of the loss imposed on producers.[71] If D is set at D^*, the optimal level of insurance given the liability load, and $r_n^* = r_f^*$, the producer's decision problem is to choose Q, s, and r to maximize $E(V)$, subject to $E(U) > U^c$ and subject to the penalty $D = D^*$ if $r < r_n^*$. But this private objective function is identical to the social welfare function; that is, it is a discontinuous function equal to equation (A1) for $r > r_n^*$ and equal to equation (A5) for $r < r_n^*$. If either $\lambda > \gamma$ or

[71] Note that this discontinuity occurs in the social (and private) welfare function for reasons related but not identical to those noted in Robert D. Cooter, Economic Analysis of Punitive Damages, 56 S. Cal. L. Rev. 79 (1982). Here markets force producers to internalize injuries to victims, so these costs are not avoided by being nonnegligent. However, the defendant's personal costs of suit are avoided.

$U_0 < U_1$, this negligence rule will lead producers to choose $r = r_n^*$, that is, to be nonnegligent.[72]

Imperfect Information. If consumers underestimate risks, under first-party liability they buy too little insurance and nonoptimal safety. Spence shows that under strict liability, a first-best solution with respect to compensation and prevention can be achieved by means of a two-part penalty.[73] A compensatory award equal to D^* is paid to victims. A fine, paid to the state initially but refunded as a subsidy to the hazardous product, is set equal to $(1 - h') (V_1 - V_0)/\overline{V}'$, where $h(p)$ is the consumer's perception of p and $(V_1 - V_0)/\overline{V}'$ is the dollar measure of utility loss implied by the consumer's willingness to pay for injury reduction.

Under a negligence rule with a risk-averse defendant and incomplete insurance, provided the standard of care is correctly set at r_f^*, the fine necessary to achieve compliance is less than under strict liability because of the discontinuity of the pay-off function. The physician will choose to meet the standard, provided

$$\frac{U_1 - U_0}{\overline{U}'} + \lambda Q \geq \frac{V_1 - V_0}{\overline{V}'} \left(1 - \frac{h'}{p'}\right) - \gamma M.$$

Thus if the load on liability insurance is at least as great as the load on first-party insurance ($\lambda > \gamma$), a fine over and above the compensatory award paid to victims is not necessary to induce compliance with a negligence standard, if the uncompensated cost of suit to defendants $(U_1 - U_0)/\overline{U}'$ exceeds the distortion in market incentives due to consumer misperceptions $[(V_1 - V_0)/\overline{V}'][1 - (h'/p')]$. Since the fine-subsidy mechanism is presumably costly to administer, this likelihood that incentives are adequate without a fine is an added attraction of a negligence rule over strict liability.

APPENDIX B

EFFECTS OF LONG STATUTES OF LIMITATIONS ON INSURANCE RISK

STANDARD MODEL OF "RISK-FREE" INSURANCE

For each of n policyholders, the expected claims cost, $E(C_i)$, is an independent drawing from a loss distribution with known mean μ and standard deviation σ. By the central limit theorem (CLT), regardless of the shape of the underlying loss distribution, the cost of the total insurance portfolio is normally distributed with mean $n\mu$ and variance $n\sigma^2$. The relative risk of the insurance portfolio, defined as

---

[72] If $\lambda > \gamma > 1$, it might be optimal to provide compensation through first-party coverage and impose a liability fine on physicians to achieve optimal deterrence. Enforcement would depend on subrogation actions by the patient's first-party insurer against the tortfeasor or his liability insurer.

[73] Spence, *supra* note 4.

the coefficient of variation, is $\sigma/\sqrt{n}\mu$.[74] As n approaches infinity, the limiting distribution of the random variable Z, defined as

$$Z = \frac{\sum_{i=i}^{n} C_i - n\mu}{\sqrt{n}\sigma},$$

approaches a standard normal distribution. Since

$$\mathrm{pr}\left(\sum_{i=i}^{n} C_i \le n\mu + Z_r \sqrt{n}\sigma\right) = r, \tag{B1}$$

where Z_r is a standard normal ordinate for confidence level r, if the firm establishes a capital reserve of $Z_r \sqrt{n}\sigma$, the probability of insolvency is $1 - r$. For example, if the desired probability of insolvency is .01, $Z_r = 2.33$.

Assume that all capital must be raised from policyholders, as in a mutual. A premium rate, R, set equal to the expected claims cost per policyholder, μ, plus a markup of 2.33 $\sigma/\sqrt{n}$, will cover losses with .99 probability. As n tends to infinity, the standard deviation of the mean loss per policyholder, $\sigma/\sqrt{n}$, and hence the capital contribution necessary to assure the desired probability of solvency, tends to zero.[75]

UNKNOWN PARAMETERS: STATIONARY DISTRIBUTION

When the parameters of the underlying loss distribution are not known with certainty, using estimates M_e and S_e in place of μ and σ in equation (B1) will result in a less than .99 probability of solvency.[76] If the underlying loss distribution is stationary, the parameters can be estimated from past experience, but the confidence interval formula must include an adjustment for sampling variation in the parameter estimates. Assuming independence from year to year and an equal number of policyholders in the estimation period (n_e) and the future policy year (n_f), the required rate is

$$R = M_e + Z_r\sigma \sqrt{\frac{1}{n_e} + \frac{1}{n_f}} = M_e + Z_r \sqrt{2}\sigma/\sqrt{n}.$$

Since $\sqrt{2} = 1.41$, adjustment for sampling error in the mean increases the safety mark-up by 41 percent, but the total still tends to zero as n tends to infinity.[77]

---

[74] J. David Cummins, Insurer's Risk: A Restatement, 41 J. Risk & Insurance 147 (1974).

[75] If n is insufficient to justify using the limiting normal distribution, then Chebyshev's inequality can be used. This states that if $\bar{X}_n$ is the mean of a random sample of size n from a distribution with mean μ and variance σ^2, then $\mathrm{pr}[|\bar{X}_n - \mu| \ge k\sigma/\sqrt{n}] < 1/k^2 = r$, where $k = r\sqrt{n}/\sigma$. Thus if $r = .01$, the required capital contribution is ten standard deviations, instead of 2.33 under normality. For use of the normal power approximation, see Venezian, *supra* note 67.

[76] Gary M. Andrews, A Note on the Use of Statistics in Rate Determination, 35 J. Risk & Insurance 320 (1968).

[77] Adjustment for unknown σ when losses are normally distributed is discussed in Cum-

Nonstationary Distribution

1. Sociolegal Risk

For the ith policyholder, claims cost for a given policy year J is the sum of claims filed in each development year t until the statute of limitations has run, T: $C_{ij} = \Sigma_{t=j}^{T} c_{it}$. (Discounting and delay in settlement are ignored and policy year subscripts are dropped hereafter except where necessary.) Assume that c_{it} is drawn from a distribution with mean μ_t and variance σ_{et}^2. The total variance, σ_{et}^2, has two components: a policy year component, σ_u^2, reflecting policyholder characteristics, and a development year component, σ_{wt}^2, reflecting legal factors. Thus expected loss for the ith policyholder is

$$E(C_i) = E\left[\sum_{t=j}^{T} (\mu_{it} + w_{it}) + u_i \right]$$

$$= \sum_{t=j}^{T} \mu_{it} = \bar{\mu}_i.$$

Expected loss on the insurance portfolio of n policies is $E\left(\Sigma_{i=1}^{n} C_i\right) = n\bar{\mu}$ with variance:

$$\text{var}\left(\sum_{i=1}^{N} C_i \right) = n\sigma_u^2 + \sum_{t=j}^{T} [n\sigma_{wt}^2 + n(n - 1)\rho_w\sigma_{wt}^2], \qquad \text{(B2)}$$

where $\rho_w\sigma_{wt}$ is the covariation among policyholders in the tth development year. The standard deviation of loss per policyholder is

$$\sigma_e = S(\bar{C}) = \sqrt{ \sigma_u^2/n + \sum_{t=j}^{T} [\sigma_{wt}^2/n + \rho_w\sigma_{wt}^2 - \rho\sigma_{wt}^2/n] }, \qquad \text{(B3)}$$

with limiting value for large n and common σ_{wt} for all t:[78]

$$\lim_{n \to \infty} \sigma_e = \sqrt{ \sum_{t=j}^{T} \rho_w\sigma_{wt}^2 } = \sqrt{T\rho_w}\,\sigma_w.$$

Thus the sociolegal risk due to development year shocks which affect all policyholders is a multiplicative function of the magnitude of the shock (σ_w), the number of development years (T), and the correlation among policyholders (ρ_w). It is independent of the size of the risk pool, n.

---

mins, *supra* note 74. For personal injury lines where the normality assumption is clearly inappropriate, more complex methods are needed.

[78] Sociolegal risk is more or less than proportional to the statute of limitations, as $\sigma_{wt} \gtrless \sigma_{wt-1}$, that is, development-year shocks increase or decrease with years after the policy year. Although number of claims filed is likely to decrease with time, size of loss is likely to increase, both because of inflation of awards and because the added cost due to decay of evidence over time eliminates relatively more minor claims.

2. *Parameter Uncertainty*

To predict the mean μ_j for policy year j, a common procedure is (1) to "develop" the immature experience of recent policy years, extrapolating from reporting patterns on prior years; (2) to fit a trend, g, through K developed previous policy years; and (3) to apply this trend to the most recent policy year (or a weighted average of the two or three most recent). Thus, $\hat{\mu}_j = \hat{g}\hat{C}_{j-1}$. The actual value is $C_j = g\mu_{j-1} + v_j + e_j$, where v is a stochastic component in the trend, with $E(v) = 0$, $E(v^2) = \sigma_v^2$ and $E(v_j, v_{j-1}) = 0$; and $e_j = u_j + w_j$ has mean zero and standard deviation σ_e given by equation (B3).

The prediction base, $\hat{C}_{j-1}$, is measured with error, $\hat{C}_{j-1} - \mu_{j-1} = e_j + s_j$, where $e_j = u_j + w_j$, and s_j is development error due to estimating ultimate experience on the basis of reporting patterns against earlier years. Assume that this development error has zero mean and is not correlated serially or with other error components: $E(s) = 0$, $E(s_j, s_{j-1}) = 0$, $E(s, v) = 0$. The discrepancy between predicted and actual value is $d = \hat{\mu}_j - C_j = (\hat{g} - g)\hat{C}_{j-1} + g(\hat{C}_{j-1} - \mu_{j-1}) - v - e$, with $E(d) = 0$, and

$$\text{var}(d) = \sigma_v^2 [1 + \hat{C}_{j-1}' (X'X)^{-1}\hat{C}_{j-1}] + (g + 1)\sigma_e^2 + g\sigma_s^2$$
$$= \sigma_v^2 + \sigma_g^2 + g(\sigma_e^2 + \sigma_s^2) + \sigma_e^2.$$

The prediction variance thus has four components. σ_v^2 is due to unexplained residual variance in the model predicting the trend. σ_g^2 is due to error in estimating the trend, g. The third term, $g(\sigma_e^2 + \sigma_s^2)$ reflects measurement error in the base used for extrapolation, $\hat{C}_{j-1}$. And σ_e^2 is the policy year residual.

Thus assuming normality, the premium rate required for a $(1 - r)$ probability of insolvency is $R = \hat{\mu} + Z_r \sqrt{\sigma_v^2 + \sigma_g^2 + g\sigma_s^2 + (g + 1)\sigma_e^2}$, which has limiting value, as $n \to \infty$, $R = \hat{\mu} + Z_r\sqrt{\sigma_v^2 + \sigma_g^2 + g\sigma_s^2 + (g + 1)T\rho_w\sigma_w}$. Long statutes of limitations contribute to nondiversifiable insurer risk through projection error (σ_g^2 and σ_s^2) and by increasing the weight attached to sociolegal changes correlated across insureds ($\sqrt{T\rho_w}\sigma_w$).