EMPIRICAL STUDY OF THE CIVIL JUSTICE SYSTEM

DANIEL P. KESSLER
Graduate School of Business, Stanford University, Hoover Institution, and National Bureau of Economic Research

DANIEL L. RUBINFELD
School of Law and Department of Economics, University of California, Berkeley, and National Bureau of Economic Research

Contents

1. Introduction 345
2. Contract law 346
   2.1. Introduction 346
   2.2. The effects of contract law versus non-contractual relations 347
   2.3. The effects of voluntary, or default rules of contract law 348
   2.4. The effects of mandatory rules of contract law 350
3. Torts 351
   3.1. Introduction 351
   3.2. Medical malpractice 352
      3.2.1. Effects of malpractice law on malpractice pressure 355
      3.2.2. Effect of malpractice law and pressure on the components of economic efficiency 356
   3.3. Auto accidents 358
   3.4. Products liability 361
      3.4.1. Effects of products liability law on liability pressure 362
      3.4.2. Effects of products liability law and liability pressure on economic efficiency 363
   3.5. Punitive damages 363
      3.5.1. Effects of punitive damages on liability pressure 364
      3.5.2. Effect of punitive damages on economic efficiency 366
   3.6. Mass torts 366
4. Property 367
   4.1. Introduction 367
   4.2. The development of private property 368
      4.2.1. Creating property rights 368
      4.2.2. Creating individual rights in shared property 369
4.3. Shared ownership 370
4.4. Property rights versus liability rules 370
4.5. Consequences for efficiency and economic growth of poorly defined/enforced property rights 372

5. The litigation process 373
5.1. Introduction: the costs of litigation 373
5.2. The chronology of a legal suit 374
5.3. Forum and law choice 375
5.4. Pre-trial 378
5.5. Settlement 381
5.6. Trial 383
5.7. Appeal 386
5.8. Alternatives to litigation: arbitration and mediation 387
5.9. Fee-shifting 389
5.10. Class actions 390
Acknowledgements 390
References 391

Abstract

In this essay, we discuss empirical research on the economic effects of the civil justice system. We discuss research on the effects of three substantive bodies of law—contracts, torts, and property—and research on the effects of the litigation process. We begin with a review of studies of aggregate empirical trends and the important issues involving contracts and torts, both positive and normative. We survey some of the more interesting empirical issues, and we conclude with some suggestions for future work. Because studies involving property law are so divergent, there is no simple description of aggregates that adequately characterizes the subject. In its place, we offer an overview of a number of the most important issues of interest. We describe (selectively) the current state of empirical knowledge, and offer some suggestions for future work. The section on legal process builds on the previous substantive sections. With respect each of the steps, from violation to trial to appeal, we review some of the more important empirical contributions.

Keywords

civil justice, civil litigation, empirical, contract, tort, property, litigation

JEL classification: K11, K12, K13, K41
1. Introduction

In this essay, we discuss empirical research on the civil justice system. Sections 2–4 discuss research on the effects of three substantive bodies of law: contracts, torts, and property. Section 5 discusses research on the effects of the litigation process. Section 5 is closely linked to sections 2–4 because each of the substantive topics (as well as other substantive bodies of law) relies on litigation for enforcement.

Research in all of the areas spans a spectrum from strongly normative to purely descriptive.

Normative economic analyses seek to evaluate the extent to which a law or choice of law apportions costs and benefits to maximize economic efficiency, defined as social surplus (consumer plus producer surplus) less transaction costs. Efficient laws appropriately balance the value of deterrence with the risk-bearing, transactions, and other costs of achieving optimal deterrence. Although it does not fit strictly in an economic framework, normative empirical legal research also examines legal systems' performance on other goals, such as distribution and corrective justice. We discuss this work below as well.

The descriptive arm of the literature is, in general, better-developed than the normative arm. This is largely due to the fact that observational data on the total costs and benefits of law-induced economic or social activity are difficult to obtain. For example, it is impossible to count the total number of contracts that are formed as a result of a body of law—much less the costs and benefits of those contracts to the parties. A typical descriptive study might report trends over time in the outcomes of cases (such as the number of cases, the stage of litigation at which a case is resolved, the probability with which a case is resolved in the plaintiff's or the defendant's favor, and the payment upon resolution of the case); the correlation between case characteristics and case outcomes; or the correlation between measures of the legal environment and case outcomes.

Descriptive analyses, particularly those relating to the effect of a law or system of laws on case outcomes, nonetheless contribute to (normative) policy debates. For example, an assessment of a law's effect on the amount or structure of litigation provides a first step toward an estimate of the magnitude of a law's transactions costs. Descriptive findings about the distributional effects of a law can also be used to assess the law's performance on the overall legal system. Finally, descriptive findings show the extent to which the actual effects of laws conform to their canonical descriptions.

Each of the first two sections of the essay follows a similar format. We begin with a review of studies of aggregate empirical trends and the important issues involving contracts and torts, both positive and normative. We survey some of the more interesting empirical issues, and we conclude with some suggestions for future work. Because studies involving property law are so divergent, there is no simple description of aggregates that adequately characterizes the subject. In its place, we offer an overview of a number of the most important issues of interest. We describe (selectively) the current state of empirical knowledge, and offer some suggestions for future research. The section on legal process builds on the previous substantive sections. With respect each of the
steps, from violation to trial to appeal, we review some of the more important empirical contributions.

2. Contract law

2.1. Introduction

Historically, contract disputes were the most common type of civil litigation in the United States. In the mid 19th century, contract cases accounted for the vast majority of all civil litigation (see Galanter, 2001 for an excellent summary of this research). Over the past 150 years, however, contract litigation has become a decreasingly important part of state courts' civil dockets, eclipsed largely by tort cases. Based on case filings in 17 states, the National Center for State Courts reports that by 1995, the number of tort cases exceeded the number of contract cases, although very recent rapid growth in contract cases has reversed this finding (NCSC, 2002).¹

In a series of reports, the Bureau of Justice Statistics (1996, 2000a) describes the characteristics of contract litigation in greater detail. Contract cases are resolved disproportionately in state courts, reflecting the fact that contract disputes are governed primarily by state law. In 1992, for example, 49,434 contract cases were commenced in federal district courts, as compared to 366,336 cases that were resolved in state courts in the Nation's 75 largest counties. Contract cases reach a conclusion relatively quickly. In 1992, the average time between filing and resolution for a contract case in a studied state court was 13 months, with a median time of 8.3 months; the average time between filing and resolution for a tort case in a studied state court was 19.3 months, with a median time of 13.7 months (Bureau of Justice Statistics, 1995). Relatively few contract cases are between individuals; most cases involved a business or a government institution. In 1992, only 7.9 percent of cases in a studied state court were between individuals, and approximately 60 percent of cases did not involve an individual at all.

We organize our review of empirical research in contract law doctrinally. The first set of studies assesses the effects of contract law versus non-contractual relations on the formation, interpretation, and enforcement of agreements. The second and third sets of studies assess the effects of particular terms of statutory contract law. We group these studies by whether they assess the effects of voluntary terms or mandatory terms (see Ayres and Gertner, 1989, but see the excellent review of empirical contract studies by Korobkin, 2002 for an alternative summary of the literature). Voluntary, or default, rules of contract fill in the gaps in incomplete contracts; parties to a contract are free to contract around them if they wish. Mandatory, or immutable, rules of contract specify the conditions necessary for a contract to be enforceable; they cannot be waived by the parties.

¹ This trend in the mix of litigated cases may not represent the trend in the mix of legal claims, if contract disputes have become disproportionately more likely to be resolved through settlement or arbitration.
2.2. The effects of contract law versus non-contractual relations

The importance of non-contractual relations in business relationships was recognized early by Macaulay (1963). In his study, Macaulay interviewed 68 businesspeople representing 43 companies and six law firms in an open format. He reported strikingly little reversion to formal contract law for the settlement of disputes over contingencies in business agreements. This was not because contingencies were absent from most agreements; his interviews suggested that contingencies were common, but that businesspeople gave little attention ex ante to how they might be resolved. Instead, he found that the value and information inherent in the ongoing relationships that characterized most of the dealings of his sample firms—combined with the high financial and non-financial costs of formal litigation—led firms frequently to resolve disputes through informal means. However, he reported that formal contract law remained important to businesspeople in some important situations: when the terms of agreements were complex; when repeat dealings were unlikely; and when the size or organization of the firm meant that employee who executed the deal needed a formal vehicle to communicate its terms to other members of the firm.

Subsequent work has largely supported Macaulay’s view. In two case studies, Bernstein (1992, 1996) examines the use by the diamond and the grain and feed industries of extralegal mechanisms for resolving contract disputes. In her earlier work, she identifies which characteristics of contract law and the diamond industry lead diamond merchants to supplant formal adjudication with their own private arbitration system. She finds that formal contract law is particularly costly to diamond merchants in each of the three efficiency terms discussed above. Uncertainty and delay in recovery increases transaction costs and the costs of the risk of losses; and the use of expectation damages lead to inefficiently too much breach, and therefore too little economic activity ex ante. On the other hand, the small number of traders and powerful reputation effects makes the alternative of relational contracting in this industry particularly low-cost. In her later work, Bernstein examines the private system of adjudication used by the National Grain and Feed Association (NGFA), a trade association of firms and individuals who are active in cash markets for grain and feed. She reports that NGFA arbitrators, as compared to generalist courts, place more weight on the specific terms of parties’ agreements than on business norms or customs. She concludes that firms prefer this private system not because non-contractual relations are unimportant to the enforcement of agreements in general, but rather because disputes that require a third party for their resolution are those in which there is little value in the parties’ ongoing relationship. Based on a comparable case-study methodology, Lorenz (1999) makes a similar argument: that a drive to enhance efficiency (and therefore gains from trade) leads firms to develop specialized procedural mechanisms to resolve disputes over contracting contingencies.

Using data from a 1995–97 survey of manufacturing firms in Ho Chi Minh City and Hanoi, McMillan and Woodruff (1999a, 1999b) document how firms in a country with poorly developed contract law use extralegal mechanisms. They find that such firms are more likely to offer credit to several types of customers: those with contracts of longer
duration, those who are members of "business networks" (through which defaulters can be sanctioned). They also find that third-party intermediaries and ethnic ties are not important mechanisms. Similarly, based on a survey of 36 Bulgarian firms in 1994, Koford and Miller (2006) report that ongoing relationships are important because of the absence of operational contract law in that country.

Johnston (1996) emphasizes how courts have come to respect these non-contractual relationships and the benefits they create. Based on a sample of 25 opinions reported by the UCC reporting service that involved a litigated issue surrounding the statute of frauds—the requirement that contracts involving the sale of goods exceeding $500 must, with certain exceptions, be in writing to be enforceable—he reports two key findings. First, he reports a strong negative correlation of the existence of a prior relationship between the parties and complete absence of any written agreement. Second, he reports a strong positive correlation of a prior relationship and the judge in the case holding that an exception to the statute of frauds applied.

Recent studies use international data to assess the efficiency effects of contract law versus non-contractual relationships. Djankov et al. (2003) measure the impact of features of 109 countries' formal contract law on likely determinants of both economic efficiency—speed of claims resolution and risk—and of fairness in the resolution of two common types of contract disputes: eviction of a tenant for non-payment of rent and collection of a bounced check. They find that increased procedural formalism—as measured by the use of professional judges and lawyers, and the restrictiveness of evidentiary and legal rules—in a country's law reduces both efficiency and fairness. Johnson et al. (2002), based on surveys undertaken in Poland, Romania, Russia, Slovakia, and Ukraine in 1997, conclude that the effect of formal contract law is important to the development of new relationships among firms and to relationships where "lock-in" is inherently low. Similarly, Bigsten et al. (2000) show that access to formal contract law in six African countries enables firms to engage in more complex transactions. Finally, Kranton and Swany (1999) document an interesting unintended adverse consequence of contract law: reductions in economic efficiency from crowding out of non-contractual relationships. Based on an historical examination of civil courts in colonial India, the authors report that the introduction of formal contract law enhanced competition in credit markets, leading not only to lower interest rates and greater efficiency, but also to unwillingness of lenders to insure farmers against adverse natural events, leading to potentially lower efficiency.

2.3. The effects of voluntary, or default rules of contract law

Voluntary, or default, rules of contract fill in the gaps in incomplete contracts. Theoretically, default rules only matter to the extent that the Coase theorem does not hold. If transaction costs are sufficiently small, and there are no bargaining costs or other market failures, then parties will contract around default rules to establish whatever terms are optimal for their particular situation. If the Coase theorem does not hold, then contract law should set default rules according to what the parties would have wanted or
intended, were the transaction costs of negotiating the omitted terms sufficiently small (Ayres and Gertner, 1989). Empirical studies of the impact of default rules generally follow the theoretical path above. The studies begin with a test of the Coase theorem. Then, to the extent that the Coase theorem is rejected, some of the studies investigate the source of the failure of the Coase theorem, and the efficiency or other implications of a particular choice of default rule.

The literature is divided on the importance of default rules. Some work finds that, at least to a first approximation, the Coase theorem holds in bargaining over contract terms. In an experimental study, Schwab (1988) found that changes in the default terms in a hypothetical labor/management negotiation exercise did not significantly prevent the parties from designing the contract in a way that maximized the sum of their rewards. Although the generalizability of this experiment to actual contract negotiations has been questioned (see Korobkin, 1998, discussed below), other work has found that parties “contract around” default terms in real-world situations. Based on a survey of the general counsels of 182 corporations on a broad range of contracting practices, Weintraub (1992) concludes that firms are able to undo default rules that are not jointly optimal. He finds that three default rules of canonical contract law—failure to enforce promptly repudiated promises, promises to deliver a good when no comparable good is available at the same price, and promises that would, due to unforeseen events, require the liquidation of the promisor—are only selectively followed in practice. Based on telephone interviews with 25 firms involved in the purchase and sale of goods, Keating (2000) investigates the impact of the default rules specified in Section 2-207 of the Uniform Commercial Code, also known as the “battle of forms” provision. (Section 2-207 governs which contract terms will govern an agreement when parties’ offer and acceptance contain conflicting, often standard-form, terms.) He finds that firms have adjusted “fairly well” to the default rules imposed by the Section, and that they “don’t seem to worry that much” about it.

Other work finds that the Coase theorem does not hold. In an experimental setting similar to Schwab’s, Korobkin (1998) found that changes in the default terms in a hypothetical negotiation over shipping services did affect parties’ contracts. Because transaction costs in the exercise were low and bargaining power was equally distributed, he hypothesizes that default terms affect contracting behavior because of “status quo bias,” due to loss aversion or regret avoidance. McCchesney (1999) investigates the effects of the rule of tortious interference—which imposes tort liability on third parties who induce a promisor to breach. In a model without transaction costs, imposing liability for tortious interference would be inefficient, since the expectation damages imposed on a promisor would be sufficient to provide incentives to breach only when it maximized the joint value of the parties relationship. But if transaction costs make renegotiation of agreements prohibitively costly, or lead damages to be systematically below their (economically efficient) expectation levels, then tortious interference may be efficient. Based on an analysis of 134 cases from the early 20th century alleging tortious interference, McCchesney finds that the doctrine affects incentives to breach, and that it is more likely to be imposed in those circumstances in which it enhances effi-
ciency. Based on interviews with construction lenders and general contractors, Mann (1996) assesses the consequences of the default rule of first-in-time creditor priority in the context of construction contracts. He compares the responses of interviewees with experience contracting in Missouri, which imposes an unusual contractor-first creditor priority rule, to the response of interviewees with experience contracting under the traditional default rule, and concludes both that the default rule affects contracting outcomes, and that the contractor-first priority rule offers efficiency advantages. Kahan and Klausner (1997) propose a novel reason why default terms might matter: increasing returns to scale, or network effects, in the use of a given default term. They study how event-risk covenants in 101 investment-grade bond issues varied between 1988 and 1993. They find that learning and network effects exist in corporate contracts, and that these externalities in general can lead privately optimal agreements to be economically inefficient from a social perspective.

2.4. The effects of mandatory rules of contract law

In contrast, mandatory rules of contract law cannot be waived by the parties. In the absence of a market failure, such rules necessarily reduce economic efficiency, because they prevent parties from agreeing on terms that they find may jointly beneficial. However, in the presence of market failures (such as asymmetric information), mandatory rules may be efficient; mandatory rules may serve other (distributive) goals as well.

Empirical studies of the effects of mandatory rules fall into two groups. The first group is largely positive in nature and focused on the effects of the doctrine of promissory estoppel. To be enforceable under canonical contract law, the promisor in an agreement had to receive some consideration in exchange for his promise; under this rule, gratuitous promises were not considered valid. The doctrine of promissory estoppel limited the application of this mandatory rule, making gratuitous promises enforceable when the promisee took an action in reliance on the promise.

The empirical literature disagrees about the importance of promissory estoppel. In an analysis of 222 cases alleging promissory estoppel decided between 1981 and 1985, Farber and Matheson (1985) find that requirement of reliance in the doctrine had been interpreted narrowly, thereby reducing the importance of the requirement of consideration. They suggest that modern contract law generally regards any agreement “in furtherance of economic activity” as enforceable. Other studies find that courts are generally unwilling to waive the consideration requirement in contract law. Pham (1994) reports that the doctrine is used infrequently, not as an exception to the consideration requirement but as an all-purpose qualification to a range of mandatory contract rules. In an analysis of 362 cases decided between July 1, 1994 and June 30, 1996, Hillman (1998) reaches a similar conclusion: that the requirement of reliance in the doctrine has been substantially retained, reducing the limitation it imposes on the requirement of consideration (and other) requirements of contract law.

Other studies are more explicitly normative in nature. Although these studies agree that mandatory terms of contract are important, they disagree over their implications.
Rice (1992), for example, examines 628 state supreme and appellate court decisions from 1900 to 1991, and finds that the application of one class of mandatory contract terms—the implied covenant of good faith—promotes neither economic efficiency nor legitimate distributitional goals of contract law. In contrast, Ayres (1995) suggests that mandatory disclosure requirements in contracting would enhance efficiency and other legitimate goals. He sent 203 testers to negotiate for the purchase of a new car according to a fixed script, varying only the race and gender of the testers. Ayres found that black and female testers paid higher prices, and concludes that mandatory disclosures by car dealers (for example, of the average price for which each make of car is sold) would lead both to more efficient markets for new cars and to reduced price discrimination on the basis of race and gender.

3. Torts

3.1. Introduction

Over the past 150 years, the number of tort law cases has expanded relative to its closest relative, contract cases. In a series of reports analogous to those on contract cases, the U.S. Department of Justice, Bureau of Justice Statistics (1995, 1997, 1999, 2000b) describes the characteristics of tort litigation in detail. Like contract disputes, tort cases are generally governed by state law and resolved in state courts. In 1994, for example, 41,166 tort cases were resolved in federal district courts, compared to 378,314 cases resolved in state courts in the Nation’s 75 largest counties in 1992. Tort cases take significantly longer to resolve than contract cases. In 1992, the average time between filing and resolution for a tort case in a studied state court was 19.3 months, with a median time of 13.7 months (Bureau of Justice Statistics, 1995). Because tort litigation primarily involves damages related to a personal injury, the vast majority of tort cases (unlike contract cases) report an individual as the plaintiff (94 percent). Tort cases disproportionately involve auto accidents. In 1992, 60.1 percent of tort cases in the Nation’s 75 largest counties involved an auto accident; the second, third, and fourth most common types of tort cases were premises liability (17.3 percent), medical malpractice (4.9 percent), and products liability (3.4 percent) (Bureau of Justice Statistics, 1995).

There is disagreement over the causes of the long-run growth in tort cases. One view, elaborated by Schwartz (1981, 1992) and Priest (1988) among others, emphasizes that the expansion of the scope of tort law is due substantially to the rise of judicial and public-policy activism in the 1960s and 1970s. According to this view, the shifting of the costs of accidental injury from (largely individual) plaintiffs to (largely institutional) defendants has occurred because of increasing acceptance of two positions: (1) institutional defendants are generally more able to prevent accidents, (2) it is good policy to provide insurance through the tort system for the costs of those accidents that can not be
prevented.\(^2\) Another view, elaborated by Galanter (1996) and Enton et al. (2000) among others, is that the growth in tort claims has been slower, more gradual, and only weakly related to reforms of the 1960s and 1970s, after appropriate adjustment for population and other factors.

Most empirical research in tort law, however, is concerned with assessment of the effects of tort law on distributional goals and economic efficiency, where efficiency (as in contract law) involves the tradeoff between transactions costs, incentives for appropriate activity, and least-cost allocation of the risk of harm or loss. We organize our review of empirical studies of tort law by type of accident, because the most important empirical issues differ for different types of accidents. Subsection 4.2 reviews the effects of tort in medical malpractice cases. The important empirical issue in this context is assessing the extent to which the cost and asymmetries of information in markets for medical care interact with conventional tort doctrine to create incentives for “defensive medicine”—the use of treatments with minimal medical benefit, or the refusal to use treatments with substantial benefit, out of fear of legal liability. Subsection 4.3 reviews the literature on the effects of tort law in auto accident cases. The most important empirical issue in this context is comparing the transaction costs, incentives for accident avoidance, and compensation paid under the traditional tort system versus a no-fault system administered through mandatory first-party insurance. Subsection 4.4 reviews work on product liability. Subsections 4.5 and 4.6 conclude with discussion of two important policy issues that span across sectors: the issue of punitive damages and the issue of mass torts. We leave the discussion of the role of the tort system in the environmental and labor law to other chapters; see Dewees, Duff, and Trebilcock (1996) for an excellent review of the empirical and theoretical issues in these contexts, and for an alternative review of the issues covered in this section. We also omit discussion of constitutional tort litigation, a federal cause of action for civil damages and injunctions against state officials who deprive others of constitutional and certain federal statutory rights; see Eisenberg and Schwab (1987) for an empirical study of this type of claim.

3.2. Medical malpractice\(^3\)

Injuries due to malpractice create substantial economic costs. According to the Institute of Medicine (2000), medical errors are the leading cause of accidental death in the United States: estimates of the number of deaths due to medical errors in 1997 range from 44,000 to 98,000. Medication errors alone account for approximately 7,000 deaths per year, exceeding the number of deaths due to workplace injuries. Weller et al. (1993) find that nearly one percent of hospital admissions in New York State in 1984 involved an injury due to negligent care; the proportion of serious injuries due to negligence was even higher.

\(^2\) But see Shavell (1987) for an analysis that shows the limited circumstances under which this hypothesis is likely to be true.

\(^3\) Pieces of this section have appeared previously in Kessler and McClellan (1996).
In theory, the liability system should provide physicians with the incentives to take the socially optimal amount of care against medical injury. By making doctors responsible for the costs of injuries that they negligently cause, the negligence rule should lead doctors' private decisions as to whether and how to practice medicine to reflect society's overall interests by leading them to balance in expectation terms the social benefits of medical care, the costs of precaution, and the costs of negligence.

In practice, several factors—most notably, the pervasiveness of insurance in the healthcare sector—drives a wedge between these seemingly-sensible incentives and socially optimal medical decision-making. On one hand, market failures may create incentives for too little precaution. First, because of transaction costs or asymmetries of information, patients often do not file malpractice claims even when there is evidence of a negligent medical injury (Harvard Medical Practice Study, 1990). Second, because malpractice liability insurance is at most weakly experience-rated (Sloan, 1990), physicians may bear little of the costs of patient injuries from malpractice. (However, physicians bear significant uninsured, nonfinancial costs of patient injuries, including the value of lost time and emotional energy in responding to a malpractice claims, Office of Technology Assessment, 1993).

On the other hand, market failures can create incentives for too much precaution, or "defensive medicine." The practice of "defensive medicine" can take two forms: positive defensive medicine and negative defensive medicine. Positive defensive medicine involves the supply of care that is relatively unproductive for patients; negative defensive medicine involves declining to supply care that is relatively productive for patients.

Positive defensive medicine is driven by moral hazard from health insurance, which means that neither patients nor physicians bear the most of the costs of care in any particular case. The costs of precautionary services financed through health insurance are generally larger than the uninsured cost of the physician's own effort. Doctors and patients make decisions that balance the costs of precaution that they bear against the costs imposed on them by the malpractice system. Thus, even if medical malpractice tort law allocated the burden of medical injuries with neither errors nor transaction costs, insensitivity to the true costs of care would lead physicians and their patients to prefer to take socially excessive precautions against injuries. The added burden due to errors and transaction costs only intensifies this effect. As Danzon (1991) observes, the fact health care providers' precautionary behavior may be ex post difficult to verify may give them the incentive to take inefficiently too much care. Craswell and Calfee (1986) prove that such excessive care results from the all-or-nothing nature of the liability decision: small increases in precaution above the optimal level can result in large decreases in expected liability.

Negative defensive medicine is driven by the fact that patients reap substantial surplus from medical care for which they can not compensate providers, while providers bear malpractice risk for which they can not charge patients. If doctors weigh the mal-
practice downside of a course of care against only a fraction of the upside, then they may withhold treatments that may be in patients' best interests.

Both forms of defensive medicine may be much more economically important than the costs of awards and settlements imposed by the malpractice system. If the costs of precaution borne by patients and physicians account for a small share of the total costs, but the costs imposed by the malpractice system are roughly proportional to the costs of awards and settlement payments, then a given change in malpractice awards would induce a much greater change in the costs of medical care. Thus, although the direct cost of malpractice tort awards, including transaction costs, is relatively insignificant at roughly one percent of medical spending (Danzon, 2000), the indirect costs of the malpractice tort system, in terms of wasteful medical treatment or loss of life due to medical errors, is likely to be far greater.

We organize our review of empirical assessment of the effects of malpractice tort law in three parts (but see Danzon, 2000 for an excellent and comprehensive alternative review). The first arm of the literature surveys physicians about their opinion of the role of the malpractice system in determining medical treatments (Reynolds et al., 1987; Moser and Musaccio, 1991; Office of Technology Assessment, 1994; Klingman et al., 1996). Although surveys indicate that physicians believe that the existing malpractice system leads to economically inefficient medical care decisions, surveys only provide information about physicians' self-reported perceptions of the effects of law: they do not measure behavior in real situations.

We discuss the remainder of the literature in two subsections below. The second arm of the literature estimates the effects of the characteristics of claims and of malpractice laws on compensation and other dimensions of "malpractice pressure"—the incentives for hospitals and physicians to shield themselves against the costs of legal liability. Although estimates from this literature have contributed to our understanding about how malpractice law works in practice, they only show how the malpractice system affects incentives, not how it affects treatment behavior—and therefore the cost and quality of care. The third arm of the literature seeks to directly measure the effects of variation in malpractice law and malpractice pressure across states on actual medical care decisions, health care costs, and patient health outcomes. This literature finds that, at least in areas with high levels of malpractice pressure, reductions in liability lead to increases in economic efficiency—that the tort system on net creates incentives for defensive medicine—although studies differ in their assessment of the magnitude of this effect. Because of the lack of empirical research, we omit discussion of two important alternatives to the tort system for apportioning damages from medical injury: no-fault insurance (but see Studdert et al., 1997; Studdert, Brennan, and Thomas, 2000; and Studdert and Brennan, 2001 for a simulation of the costs and benefits of a hypothetical no-fault system) and contract (which, because of the pre-existing relationship between the parties, is a viable alternative to tort (e.g., Shavell, 1987); see Havighurst, 1995 for a theoretical evaluation of apportionment of medical liability through contract).
3.2.1. Effects of malpractice law on malpractice pressure

Simple statistical indicators suggest that the role of the malpractice system in health care has grown over the past 40 years. As Danzon (2000) points out, the number of malpractice claims per physician and the award paid per claim increased rapidly in the US from the 1960s to the 1980s. Claim frequency increased at more than 10 percent per year, reaching a peak of 17 claims per 100 physicians in 1986. Awards paid per claim increased at roughly twice the rate of inflation (Danzon, 1986), with some evidence of even greater growth for the most costly cases (Shanley and Peterson, 1987). In turn, malpractice liability insurance markets experienced two “crises,” one in the mid-1970s and one in the mid-1980s, in which prices for malpractice insurance skyrocketed and availability of insurance contracted (see Danzon, 2000, section 4.2, for a cataloguing of the literature on malpractice insurance). Taken together, these factors led states’ legislatures to change their laws governing medical malpractice claims to reduce liability.

Numerous studies have investigated the effects of the characteristics of malpractice claims and tort law on measures of malpractice pressure such as malpractice claims rates, the awards paid to and other measures of the disposition of claims, and malpractice liability insurance premiums. This literature reports three main findings. First, economic loss, rather than fault, is consistently the most important characteristic of claims in determining the probability and size of award (Danzon and Lillard, 1983; Farber and White, 1991; Brennan et al., 1996). Second, the changes in state laws designed to reduce liability significantly reduced the incentives to supply preventative care. Danzon (1982, 1986) and Sloan et al. (1989) find that tort reforms that cap physicians’ liability at some maximum level or require awards in malpractice cases to be offset by the amount of compensation received by patients from collateral sources reduce payments per claim. For example, based on 1975–1978 data, Danzon (1982) reports that states enacting caps on damages had 19 percent lower awards, and states enacting mandatory collateral source offsets had 50 percent lower awards. Based on 1975–1984 data, Danzon (1986) reports that states enacting caps had 23 percent lower awards, and states enacting collateral source offsets had 11 to 18 percent lower awards. Based on 1975–1978 and 1984 data, Sloan et al. (1989) find that caps reduced awards by 38 to 39 percent, and collateral source offsets reduced awards by 21 percent. Danzon (1986)

---

4 Reforms requiring collateral-source offset revoke the common-law default rule which states that the defendant must bear the full cost of the injury suffered by the plaintiff, even if the plaintiff were compensated for all or part of the cost by an independent or “collateral” source. Under the common-law default rule, defendants liable for medical malpractice always bear the cost of treating a patient for medical injuries resulting from the malpractice, even if the treatment were financed by the patient’s own health insurance. Either the plaintiff enjoys double recovery (the plaintiff recovers from the defendant and his own health insurance for medical expenses attributable to the injury) or the defendant reimburses the plaintiff’s (subrogee health insurer, depending on the plaintiff’s insurance contract and state or federal law. However, some states have enacted reforms that specify that total damages payable in a malpractice tort are to be reduced by all or part of the value of collateral source payments.
also finds that collateral-source-rule reforms and statute-of-limitations reductions reduce claim frequency. Based on data from malpractice insurance markets, Zuckerman et al. (1990) and Barker (1992) reach similar conclusions: Zuckerman et al. find that caps on damages and statute-of-limitations reductions reduce malpractice premiums, and Barker finds that caps on damages increase profitability. Third, the two reforms most commonly found to reduce payments to and the frequency of claims, caps on damages and collateral source rule reforms, share a common property: they directly reduce expected malpractice awards. Caps on damages truncate the distribution of awards; mandatory collateral source offsets shift down its mean. Other malpractice reforms that only affect malpractice awards indirectly, such as reforms imposing mandatory periodic payments (which require damages in certain cases to be disbursed in the form of annuity that pays out over time), have had a less discernable impact on liability and hence on malpractice pressure.

Taken alone, estimates of the impact of reforms on frequency and severity from these analyses are only the first step toward answering the policy question of interest. They provide evidence only of the effects of legal reforms on doctors’ incentives; they do not provide evidence of the effects of legal reforms on doctors’ behavior. Identifying the economic efficiency of precautionary behavior due to legal liability requires a comparison of the response of costs of precaution and the response of losses from adverse events to changes in the legal environment.

3.2.2. Effect of malpractice law and pressure on the components of economic efficiency

The third arm of the literature investigates how treatment decisions and patient health outcomes respond to malpractice pressure. Early work estimates the effect of physicians’ actual exposure to malpractice claims to clinical practices and outcomes (Rock, 1988; Harvard Medical Practice Study, 1990; Localio et al., 1993; Baldwin et al., 1995). Rock, Localio et al., and the Harvard Medical Practice Study find results consistent with defensive medicine; Baldwin et al. do not. However, concerns about unoberved heterogeneity across providers and across small geographic areas qualify the results of all of these studies. These studies use frequency of claims or magnitude of insurance premiums at the level of individual doctors, hospitals, or areas within a single state over a limited time period to measure malpractice pressure. Because malpractice laws within a state at a given time are constant, the measures of malpractice pressure used in these studies arose not from laws but from primarily unobserved factors at the level of individual providers or small areas, creating a potentially serious problem of selection bias. For example, the claims frequency or insurance premiums of a particular provider or area may be relatively high because the provider is relatively low quality, because the patients are particularly sick (and hence prone to adverse outcomes), because the patients had more “taste” for medical interventions (and hence more likely to disagree with their provider about management decisions), or because of many other factors; the sources of the variation in malpractice pressure are unclear and probably multifactor-
ial. All of these factors are extremely difficult to capture fully in observational datasets, and could lead to an apparent but non-causal association between measured malpractice pressure and treatment decisions or outcomes.

More recent work seeks to address this endogeneity problem by identifying the effects of malpractice pressure with variation across states and over time in malpractice law reforms, which are arguably exogenous (but see Danzon, 2000 for a critique of this assumption). Much of this work has investigated the consequences of malpractice pressure for positive defensive medicine. Kessler and McClellan (1996) use longitudinal data on essentially all elderly Medicare beneficiaries hospitalized with serious cardiac illness from 1984, 1987, and 1990, matched with information on the existence of direct and indirect law reforms from the state in which the patient was treated. They found that reforms that directly limit liability—such as caps on damages—reduced hospital expenditures by 5 to 9 percent in the late 1980s, with effects that are greater for ischemic heart disease (IHD) than for heart attack (AMI) patients.5 In contrast, reforms that limit liability only indirectly were not associated with any substantial expenditure effects. Neither type of reforms led to any consequential differences in mortality or the occurrence of serious complications. The estimated expenditure/benefit ratio associated with liability-pressure-induced intensive treatment was over $500,000 per additional one-year survivor, with comparable ratios for recurrent AMIs and heart failure. Thus, treatment of elderly patients with heart disease does involve defensive medical practices, and limited reductions in liability can reduce this costly behavior.6

Two recent studies identify the mechanism through which direct reforms affect physician behavior, in order to help predict whether existing reforms under new market conditions, or new and untried types of reforms, will have similar effects. Kessler and McClellan (2002b) match longitudinal Medicare data with law reforms and data on health insurance markets to explore the ways in which managed care and liability reform interact to affect treatment intensity and health outcomes. They report that direct reforms reduce defensive practices in areas both with low and with high levels of managed care enrollment. Managed care and direct reforms do not have long-run interaction effects that are harmful to patient health. However, at least for patients with less severe cardiac illness, managed care and direct reforms are substitutes, so the reduction in defensive practices that can be achieved with direct reforms is smaller in areas with high managed care enrollment.

Kessler and McClellan (2002a) integrate four unique data sources to illuminate how reforms affect malpractice pressure, and how reform-induced changes in the incentives provided by the liability system affect treatment decisions, medical costs, and health outcomes. That paper matches by state and year the longitudinal Medicare data discussed above (updated to include all years from 1984–1994) with data on law reforms;

5 Because IHD is a less severe form of illness, IHD patients may have more “marginal” indications for intensive treatment, leading to a greater scope for defensive practices.

6 Dubay et al. (1999) confirm that defensive practices exist in non-elderly populations, although they report that the costs of defensive medicine in obstetrics are small.
physician-level data on the frequency of malpractice claims from the American Medical Association’s Socioeconomic Monitoring System (AMA SMS); and malpractice-claim-level data from the Physician Insurers Association of America on claim costs and claim outcomes. They report that although direct reforms improve medical productivity primarily by reducing malpractice claims rates and compensation conditional on a claim, other policies that reduce the time spent and the amount of conflict involved in defending against a claim can also reduce defensive practices substantially. For example, at least for elderly heart disease patients, an untried reform that reduced the legal-defense burden on physicians and hospitals by one-quarter—which is within the range of policy possibilities—could be expected to reduce medical treatment intensity by approximately 6 percent, but not to increase the incidence of adverse health outcomes. In the same population, a policy that expedited claim resolution by six months across-the-board could be expected to reduce hospital treatment costs by 2.8 percent, without greater adverse outcomes. This finding is consistent with Kessler and McClellan (1997), which reports broad differences in physicians’ perceptions of the impact of malpractice pressure in states with and without liability reforms.

Other work has investigated the consequences of malpractice pressure for negative defensive medicine. For example, Dubay et al. (2001) find that a decrease in malpractice premiums that would result from a feasible policy reform would lead to a decrease in the incidence of late prenatal care by between 3.0 and 5.9 percent for black women, and between 2.2 and 4.7 percent for white women. However, although they found evidence that malpractice pressure was associated with greater delay and fewer prenatal visits, they found no evidence that this negatively affected infant health. More recent work finds substantial evidence of negative defensive medicine. Hellinger and Encinosa (2003) report that the supply of physicians was approximately 12 percent greater in states with caps on non-economic damages, as compared to states without them.

3.3. Auto accidents

After medical errors, auto accidents are the second-leading cause of accidental death in the United States, with just over 40,000 deaths per year (U.S. Department of Transportation, 2003). Adding the costs of less severe injuries increases the estimated magnitude of costs of auto accidents. Roughly two-thirds of all civil litigation involves an automobile accident (Rolph et al., 1985). According to the Urban Institute (1991), auto accidents cause a total of $420 billion in total damage every year. Approximately one-quarter of this amount, or $100 billion, is passed through the automobile liability insurance system.

Empirical research in the auto accident context has focused on comparisons of the performance of tort versus no-fault systems (but see Vickrey, 1968; Wittman, 1986; White, 1989; Kessler, 1995; and Edlin, 2003 for comparisons of the incentives for precaution provided by different systems of tort law). Under auto no-fault systems, some proportion of losses from accidents are compensated by a driver's own (mandatory) insurance regardless of negligence, with the remainder apportioned by tort law. Under tort
law, losses from accidents are borne by the party at fault. The empirical question generally takes the following form: can no-fault systems achieve the same level of deterrence as tort systems, but with lower transactions costs and significantly better performance on generally-accepted compensation goals?

There is substantial agreement that no-fault systems administer compensation for auto accidents with much lower transaction costs than does the tort system. According to Carroll et al. (1991), costs of litigation, settlement, and other administration of compensation in a typical tort liability system amounted to about 33 percent of the cost of injuries covered by insurance; this excludes the publicly-financed costs of administering the civil justice system, which are substantial (Kakalik and Robyn, 1982). Under a typical no-fault plan, they find that transaction costs would be reduced by about 39 percent. On a base of $100 billion per year (Carroll et al.), this amounts to a lower-bound estimate of efficiency improvements of $13 billion per year (100*.33*.39).

Carroll et al. (1991) also show that no-fault would deliver compensation faster and make it more closely track economic loss. Under a typical tort system, claimants receive initial compensation payments 181 days after the accident; under a typical no-fault system, claimants receive initial compensation payments 116 days after the accident, or 36 percent faster. In addition, under a typical tort system, claimants with less serious injuries tend to receive more than their economic loss (see also Carroll and Abrahamse, 1999, and the work cited therein, for evidence on the extent of overcompensation due to fraud and abuse), with fully 62 percent of all injured people receiving more than their economic loss; claimants with more serious injuries tend to receive less than their economic loss, with 27 percent of injured people receiving less than their economic loss. By comparison, under a typical no-fault system, only 22 percent of injured people receive more than their economic loss, and 16 percent receive less than their economic loss.

From a theoretical perspective, the overall impact of mandatory no-fault on accident rates is ambiguous. In theory, mandatory no-fault compensation for damages unambiguously reduces the incentives to take care, because drivers do not directly bear the costs of their negligence. In practice, however, several factors might mitigate this result. Under a typical no-fault system, a driver would still have incentives to take care to avoid injury, to avoid criminal and traffic penalties, and to avoid the premium increases that would likely accompany an accident under a conventionally-experience-rated no-fault insurance policy (Loughran, 2001).

The empirical papers investigating the effects of no-fault on accidents share a common research design. They use data at the state/year level on the rate of fatal auto accidents; the characteristics of states and their populations, such as weather, alcohol consumption, real income, miles of roads, the mean and variance of vehicle speeds, the proportion of vehicle registrations that are trucks, hospitals per square mile, and the age, education, and employment statuses of their population; and state auto accident law, including whether or not the state had mandatory no-fault insurance, the scope of losses covered by the no-fault system; and other laws governing the apportionment of damages from auto accidents. They estimate the effect of no-fault on the fatal accident
rate by comparing the trend in the fatal accident rate in (all or a subset of) the seventeen states that adopted no-fault insurance between 1971 and 1990 to the accident rate in all other states, holding constant other covariates.

These papers generally find that no-fault leads to greater numbers of fatal accidents, with some papers based on earlier data finding no effect. Based on 1967 to 1975 data, Landes (1982) finds that no-fault systems with relatively low "tort thresholds" lead to increases in fatal accident rates of approximately 4 percent, and (more expensive) no-fault systems with high tort thresholds lead to increases in accident rates in excess of 10 percent. In contrast, follow-up work by Kochanowski and Young (1985) using data from 1975, 1976, and 1977, and by Zador and Lund (1986), using data from 1976 to 1980, found no effect of no-fault on accident rates.

Later work, and work based on data from other countries, has consistently found that the decreased incentives created by no-fault increases the accident rate. Cummins et al. (2001) use data from 1968 to 1994. Using simple ordinary least squares regression ("OLS") models, they find that no-fault increases the accident rate statistically significantly by 5 percent; in simultaneous-equations models of the joint process by which states adopt no-fault and the effect of no-fault, they find that no-fault increases the accident rate by as much as three times this amount. Cohen and Dehejia (2003) use data from 1970 to 1998 and find in OLS models that no-fault leads to an increase in traffic fatalities of approximately 6 percent. Sloan et al. (1995) report that tort liability accomplishes this reduction in fatal accidents by reducing the prevalence of one of the key determinants of auto accidents—binge drinking.7 McEwain (1989), using state data from Australia and New Zealand from 1970 to 1981, and Devlin (1990), using provincial data from Canada from 1967 to 1984, also find positive effects of no-fault on the fatal accident rate of 16 and 9 percent, respectively.

In a series of papers, O'Connell and various coauthors (O'Connell and Joost, 1986; O'Connell et al., 1993, 1995, 1996), Abrahamse and Carroll (1995), and Carroll and Abrahamse (1999) argue that a "choice" system of auto insurance offers a viable compromise that addresses the limitations of both mandatory no-fault and tort. Under a choice system, drivers choose whether to be insured under a traditional tort law or under an administrative no-fault system. Those who choose the tort system retain traditional tort rights and responsibilities, but those who choose no-fault can neither recover for nor are liable for those losses that are designated to the no-fault system. However, as Kabler (1999) points out, none of this work explains why a choice system would avoid the adverse accident effects of no-fault systems. And, adverse selection by (high-premium) risky drivers into the no-fault system may decrease the value of retaining tort coverage so much that a choice system would collapse into mandatory no-fault.

---

7 See Chaloupka et al. (1993), Sloan et al. (1994), and Rahn (1996) for a comparison of the effectiveness of tort liability versus other approaches for deterring fatalities due to drinking and driving.
3.4. Products liability

Losses from injuries from products are controlled differently than are losses from other types of accidents. First, regulatory mechanisms other than tort provide important incentives for product safety. The Consumer Products Safety Commission, The Occupational Health and Safety Administration, The Food and Drug Administration, The Federal Aviation Administration, and the Environmental Protection Agency all regulate the safety of products. Second, over the past three decades, most US jurisdictions have adopted strict liability for injuries from products (Keeton et al., 1984). This means that manufacturers and retailers are liable for damages caused by product defects, even if they exercised all possible care in the preparation and sale of the product (Restatement (Second) of Torts, Section 402A (1977)).

This combined regulatory/liability system could provide incentives for either too much or too little precaution from the perspective of economic efficiency. On one hand, the joint regulatory/liability system may provide incentives for economically efficient precautionary care decisions as well as or better than the basic “negligence rule.” For example, if regulatory agencies’ monitoring is infrequent and ability to impose sanctions is limited, and transaction costs limit the tort system’s ability to provide sufficient incentives for safety, then a combined regulatory/liability system may be socially preferred to either system standing alone. On the other hand, the combined costs imposed by a system of strict liability for product-related accidents and the regulatory system may lead to too much precaution. This excessive precaution can take two forms: socially insufficient or misdirected product innovation and insufficient or misdirected product use (e.g., activity levels).

We organize our review of the empirical literature in the same three parts as in medical malpractice (see also Litan, 1991 and Congressional Budget Office, 2003 for excellent alternative reviews). The first arm of the literature seeks to assess the impact of law and regulation using surveys and case studies. Some of this work finds product liability law has small or ambiguous effects on innovation and activity level (e.g., Eads and Reuter, 1985; Johnson, 1991); some finds evidence of adverse efficiency effects (e.g., Mackay, 1991); and some finds evidence of both socially constructive and adverse effects (e.g., Garber, 1993). However, because of the context-specific nature of this work, the findings are often difficult to generalize beyond the firms or individuals featured in the study.

We discuss the remainder of the literature in two subsections below. As in medical malpractice, we first describe work that estimates the effects of the characteristics of claims and of products liability laws on compensation and other dimensions of “liability pressure”—the incentives for product safety. Then, we discuss work that seeks to directly measure the effects of variation in products liability law and liability pressure on the three components of the efficiency costs and benefits of products liability law: accidents and deaths; the costs and availability of products; and measures of innovation, research, development, and productivity.
3.4.1. Effects of products liability law on liability pressure

Numerous studies have sought to assess whether liability pressure in the products area has been rising or falling nationally over the past 20 years. Indices constructed by Eisenberg and co-authors from filed cases conclude that liability pressure has been falling (Henderson and Eisenberg, 1990; Eisenberg and Henderson, 1992; Eisenberg et al., 1996; Eisenberg, 1999). However, these analyses do not account for liability pressure created by claims that were not litigated, i.e., that did not result in a court filing. To the extent that the selection of cases has been changing over time, indices based only on filed cases will be biased (but see Eisenberg and Henderson, 1992 for an argument why selection bias is unimportant to their findings). Indeed, indices based on insurance premiums and other data that include all payments to claimants, whether or not they result in a court case, find a dramatic increase in the cost of products liability (e.g., Viscusi, 1991). The expansion of compensation through the tort system reported in these studies is consistent with the findings of Henderson (1991), who finds “fairness” more important than “efficiency” in judges’ reasoning in products liability cases.

Although this work provides background about trends over time in the incentives provided by the tort system, it does not identify how products liability law affects incentives for care. Since many determinants of liability pressure other than law have changed over time, national trends reflect some combination of these effects and changes in law. Three studies provide evidence that limitations on liability reduce liability pressure. Viscusi (1990) uses data from products liability insurance markets from 1980 to 1984. He finds that state statutes designed to reduce liability pressure—including those that provide a state-of-the-art defense, a statute of limitations for producer liability, and collateral-source offsets and other damages restrictions—reduce products liability premiums, reduce insurer loss ratios, and increase insurance availability. Manning (1994) estimates the effect of liability costs on product prices by comparing the difference in trends from 1950–1989 in the prices of the DPT vaccine (which confers immunity from diphtheria, pertussis (whooping cough), and tetanus) and the DT vaccine (which confers immunity from diphtheria and tetanus only). He argues that this identifies the impact of liability costs because the DPT and DT vaccines have roughly comparable costs of production, but the DPT vaccine has significantly higher liability costs (because of the possibility of rare but serious complications from its pertussis component). He finds that the wholesale price of the DPT vaccine rose by over 2,000 percent during the study period, and that 96 percent of the price increase was due to litigation costs. Manning (1997) estimates the effect of liability costs on the relative U.S./Canadian prices charged by the manufacturer of 121 of the 200 most prescribed prescription drugs in 1990. He compares the relative prices of drugs with substantial liability risks, as measured by litigation history, controlled substance designation, relationship to vaccine liability costs, risk-assessment survey response, and drug-reference book risk rating, to the relative prices of drugs without such risks, holding other determinants of relative prices constant. He finds that liability risk roughly doubles the
average US/Canadian price differential, and increases the median price differential by about one-third.

3.4.2. Effects of products liability law and liability pressure on economic efficiency

The third arm of the literature investigates how products liability law affects economic efficiency. Some studies identify the efficiency effects of products liability law by comparing national trends in accident rates after versus before periods of expansion of product liability pressure. Priest (1988) compares trends in the rate of accidental deaths and injuries from products for the late 1970s to trends for the 1960s and early 1970s. He finds an increase in product deaths and injuries in the later period, and concludes that the expansion of product liability had minimal (if any) beneficial effects on accident avoidance. Martin (1991) analyzes data from the aviation industry using a similar methodology. He finds sharper declines in the aviation accident rate from 1950–69 than from 1970–89.

Like the work on trends over time in the incentives provided by the tort system, though, this work is only suggestive, since many other determinants of the accident rate (such as regulatory policy) may have been changing contemporaneously with aggregate trends in products liability pressure. Two studies attempt to control for such factors. Campbell et al. (1998) use data on labor productivity (gross state product per worker) by industry from 1970 to 1990. They compare trends in productivity in states that adopted liability law reforms, such as caps on damage awards, to trends in productivity in states that did not, holding constant fixed differences across states and the time-varying political and economic characteristics of states. They find that states that changed their liability laws to reduce levels of liability experienced greater increases in productivity that states that did not. Viscusi and Moore (1993) use data from 1980–84 from the Profit Impact of Marketing Strategies (PIMS) study, containing balance sheet and income statement items for distinct lines of business from large firms, matched by 3-digit SIC industry code with data from the same period on product liability cost measures from the Insurance Services Office. They report a positive effect of liability on product innovation at low to moderate levels of liability pressure, but a negative effect of liability on product innovation at high levels of liability. In addition, they find that liability pressure has a greater effect on product innovation than on process innovation.

3.5. Punitive damages

The economic effects of punitive damages—those in excess of what is necessary to compensate a plaintiff for his economic and noneconomic losses—is an important policy issue that applies to each of the four types of tort cases discussed above. On one hand, punitive damages may be necessary to induce optimal precautionary behavior in two situations: those in which injurers’ behavior is undetected, and those in which can escape liability, and those in which the benefits enjoyed by injurers should be excluded
from social welfare (Polinsky and Shavell, 1998). On the other hand, if punitive damages are awarded in other situations, or in amounts disproportionate to the probability of defendants’ escaping liability, then they may lead to socially excessive precaution.

We organize our review of the empirical literature in the same three parts as in medical malpractice and products liability (see also Robbennolt, 2002 for an excellent alternative review of the empirical literature on punitive damages). The first arm of the literature seeks to assess the impact of punitive damages using surveys. Baker (1998) reports that lawyers believe that statutory restrictions on punitive damages would be at best only partially effective, because juries will inflate compensatory damages to offset their reduced ability to impose punitive damages. Launrie et al. (1994) report that a sample of 100 small business owners and 100 corporate counsel believe that punitive damages increase estimated settlement payments by 13 percent. However, as in the case of surveys in products liability, the context-specific nature of this work makes findings difficult to generalize beyond the firms or individuals featured in the study.

We discuss the remainder of the literature in two subsections below. As in medical malpractice, we first discuss work that estimates the effects of punitive damages on measures of “liability pressure”—the incentives for safety. Then, we discuss work that seeks to directly measure the effects of punitive damages on the efficiency of tort law.

3.5.1. Effects of punitive damages on liability pressure

Numerous studies have sought to assess whether punitive damages have an important effect on the incentives for accident avoidance. Most of this work has investigated the frequency and magnitude of punitive damages awards in litigated cases (Peterson et al., 1987; GAO, 1989; Daniels and Martin, 1990; Hensler and Moller, 1995; Moller et al., 1997; Eisenberg et al., 1997, 2002). The studies report similar overall probabilities of a punitive damage award, ranging from 4.7 percent of jury verdicts in Cook County, Illinois and California from 1980–84 (Peterson et al., 1987), to 4.9 percent of 25,627 jury verdicts from state trial courts in eleven states from 1981–85 (Daniels and Martin, 1990), to 6 percent of 6,053 trials in the Civil Trial Court Network sample of state trial courts in 45 counties in 1991–92 (Eisenberg et al., 1997). However, punitive damages are awarded substantially more frequently in certain types of cases, including those involving financial injuries (Moller et al., 1997); business/contract cases (Peterson et al., 1987); and fraud and intentional tort cases (Peterson et al., 1987; Eisenberg et al., 2002). The median punitive damage award is roughly equal to the median compensatory damage award, but the mean punitive damage award is much larger (Eisenberg et al., 1997), because of a small number of very large punitive damage cases. The two RAND studies that examine punitive damage awards over time (Peterson et al., 1987; Hensler and Moller, 1995) find that although the frequency of punitive damage awards in trials have remained roughly constant, total punitive-damage dollars awarded has risen substantially. However, as in the products liability context, the frequency, severity, or trend in frequency or severity of punitive damages awarded in litigated cases does not account for liability pressure created by claims that were not litigated, i.e., that did not result in
a court filing. As Polinsky (1997) explains, because litigated cases are such a small and characteristically unrepresentative sample of all claims, extrapolating the results from these studies to the universe of claims is not possible without substantial additional assumptions that are unlikely to be correct.

Other studies have sought to identify the liability pressure created by punitive damage awards with broader measures. Koenig (1998) reports on several studies conducted by the Texas Department of Insurance (TDI) analyzing all closed liability insurance claims in that state. Insurance adjusters review each claim in the TDI data and apportion settlement payments into compensatory and punitive damages. In these studies, the TDI estimates the effect that statutory restrictions on punitive damages adopted by Texas in 1995, requiring plaintiffs to show clear and convincing evidence of malice to obtain punitive damages. TDI concludes that punitive damages accounted for between 10 and 16 percent of all claims payments, depending on the size of the claim and the time period and that the restrictions on punitive damages would reduce liability insurance premiums by 4.5 percent, or approximately $428 million.

As with litigated cases, the liability pressure created by punitive damages varies substantially across case types. Using 1983 data on commercial bodily injury (other than medical malpractice) insurance claims of $25,000 or more, the Insurance Services Office (1988) shows that less than one percent of payments to such claims are attributable to punitive damages. Karpoff and Lott (1999) use event studies to estimate the impact on firms’ stock prices of several key Supreme Court decisions, congressional actions, and specific lawsuits that changed firms’ exposure to punitive damages liabilities. They find little evidence that the announcement of changes in the legal environment affect valuations, but substantial evidence that the announcement of specific lawsuits affect valuations by more than the direct cost of the punitive damage award.

A new literature based on jury experiments investigates whether punitive damages are awarded in situations and in amounts that would create the incentives for optimal precaution (Baron and Ritov, 1993; Sunstein et al., 1998; Viscusi, 2001a, 2001b, 2002). The experiments ask jury-eligible participants to report the punitive damages that they would award in hypothetical fact situations under different sets of jury instructions or statutes. The experiments also ask participants about the reasoning that they used to reach their decision. The studies find that punitive damages depend neither on the defendant’s ability to escape detection (as would be necessary to achieve economic efficiency) nor on other case characteristics that would be necessary to achieve most notions of distributitional “fairness”. In addition, the studies conclude that offering juries additional guidance by increasing the specificity of jury instructions (e.g., by providing them with a mathematical formula for determining punitive damages) have little effect (see especially Viscusi, 2001b). Sunstein et al. (1998) suggest that this is due not to differences in individuals’ moral judgments or interpretations of fact situations, but rather to their difficulty in translating such judgments into financial terms.
3.5.2. Effect of punitive damages on economic efficiency

Only one study seeks to assess the effect of punitive damages on economic efficiency. Viscusi (1998) compares several measures of the costs of accidents from states that allow punitive damage awards to those from the four states that do not allow punitive damage awards (Michigan, Nebraska, New Hampshire, and Washington), controlling for other characteristics of states that may affect accident costs. He finds that allowing punitive damages has no significant effect on the rate of toxic chemical accidents, on toxic chemical releases, or on accidental death rates (but see the critique in Eisenberg, 1998). Because punitive damages have no effect on precautionary behavior, but impose substantial uncertainty and transaction costs on potential injurers, he concludes that they decrease efficiency.

3.6. Mass torts

How the liability system handles mass torts is another important policy issue that stretches across case types (see Hensler, 2001 for an excellent review and empirical analysis of mass torts, which we largely follow below). “Mass tort” is not a formal legal term; it usually denotes the consolidated litigation of a large number of tort claims arising out of a single accident or use of a single product. Such consolidation may include certification as a class action, but may also include other, less formal and less restrictive procedures (for discussion, see Peterson and Selvin, 1988). For example, “multi-districting,” one doctrine used in mass tort cases, allows federal courts to aggregate cases involving common questions of fact into a single court by assigning responsibility for managing some or all phases of the litigation to a specific judge (Hensler et al., 1985). Empirically, mass torts account for only one-fifth of class action lawsuits, and class actions account for approximately 35 percent of mass torts (Hensler, 2001). In the 1980s, mass tort litigation grew significantly; Hensler and Peterson (1993) describe in detail the cases, including litigation over Agent Orange, Bendectin, the Dalkon Shield, the Hyatt skywalk collapse, the MGM Grand hotel fire, asbestos, and others, and provides some hypotheses about its causes.

As with class actions, the main efficiency argument in favor of the various mass tort procedural techniques is that they create economies of scale in litigation, thereby reducing the costs of evaluating numerous similar claims. No empirical studies have assessed comprehensively the efficiency or distributional consequences of mass-tort treatment of claims. However, in a number of formal reports and academic papers, researchers at the RAND Institute for Civil Justice have examined the characteristics of the largest and most costly (and still ongoing) of all mass torts—the asbestos litigation. Carroll et al. (2002) contains a comprehensive description of this work and bibliography; we summarize its key findings here.

Asbestos is a naturally-occurring mineral that is inexpensive to mine and process. Asbestos has many favorable properties, including strength, durability, and fire-retardancy. However, inhalation of asbestos fibers that occurs as a result of its handling and use
causes a variety of lung diseases. Widespread occupational exposure to asbestos from 1940 to 1979 has been estimated to cause more than 225,000 premature deaths from 1985 though 2009, and large numbers of other less severe injuries and disability. After the 1973 decision in Fibreboard vs. Borel\(^8\) finding asbestos manufacturers strictly liable to workers injured as a result of their product, increasing numbers of product liability lawsuits began to flow into the courts.

Carroll et al. (2002) report that over 600,000 people have filed asbestos claims through the end of 2000, with annual filings rising sharply in recent years. Over 6,000 companies in 75 out of 83 different types of industries have been named as defendants, ranging to include essentially every type of economic activity in the U.S. They find that a total of $54 billion has already been spent on asbestos litigation, over half of which has been consumed by transaction costs. Increasing numbers of less severe injuries account for most of the recent growth in the asbestos caseload. Estimates of the number of future claims and future costs vary widely. Best case scenarios predict approximately 1.2 million claimants and a total cost of $200 billion; worst case scenarios predict 3 million claimants and a total cost of $265 billion. The RAND researchers conclude that for asbestos compensation, the tort system is falling short of all of its principal objectives: deterrence of potential injurers, compensation of injured victims, and provision of "corrective justice."

4. Property

4.1. Introduction

The empirical literature relating to private property is quite diverse in its coverage. If there is one central theme, however, it relates to the design of mechanisms for allocating property rights and the rules for transfer of those rights, in the presence of transactions costs. The principal tradeoff behind these mechanisms is between incentives for optimal investment and transaction costs. This theme is not entirely new. Property law resembles contract law in situations where transaction costs are typically low, and resembles tort law in which transaction costs are often high.

In 4.2 we review studies of the development of private property; we follow this with an analysis of the efficiency/transaction costs tradeoff of different rules for the individual acquisition of previously shared property. Section 4.3 examines a closely related issue: the efficiency consequences of different rules of shared ownership—the allocation of rights to property for which the transaction costs of creation of individual rights is prohibitively high. In 4.4 we review the work that defines what the boundary of optimal property law should be, i.e., when property rules should be forgone in favor of liability rules. Work in this area was fundamentally shaped by Ronald Coase (1960), who

\(^8\) 493 F.2d 1076, cert. denied 419 U.S. 869, 95 S.Ct 127, 42 L.Ed.2d 107.
observed that any allocation of property rights is efficient as long as there are no transaction costs and property rights are well-defined and enforced. Finally, 4.5 examines the consequences in terms of economic growth of the violation of the second assumption behind the Coase theorem: when property rights are poorly enforced and/or poorly defined.

4.2. The development of private property

The development of rights of private property has been a fundamental underpinning of our capitalist economic system. In the subsection that follows we discuss the creation of property rights generally; we follow this with a more focused discussion of property rights on shared property.

4.2.1. Creating property rights

Theoretical debates have surrounded such issues as the time span of private property rights, the alienability of those rights, and the appropriate use of public lands. Accompanying and complementing these theoretical debates have been a wide range of historical and empirical studies of the development and evolution of private property rights. For example, citing evidence from a study of Indians in Labrador, Demsetz (1967) posited the theory that all property rights regimes evolve efficiently in response to changes in economic conditions. Ellickson (1993) focuses his attention on regimes associated with close-knit societies in which power is broadly dispersed and members interact regularly. He questions the ability of Demsetz' theory to explain the development of the institution of slavery and other coercive regimes associated with cultures that were not close knit. With respect to close-knit societies only, Ellickson points to economic efficiency, to the importance of non-economic concerns such as liberty, privacy, equality, and community. In support of his broader theory, he offers case studies of land ownership regimes at settlements during the colonial period in U.S history (Jamestown and Plymouth) as well as other non-U.S settlements, including those in Israeli (kibbutzim) and Mexico (ejidos).

A number of scholars have gone further back in time in their analysis of the evolution of private property, focusing on the development of rights of aboriginal peoples. Bailey (1992) reviews a wide range of studies of aboriginal societies, concluding that economic considerations play a significant role in explaining the evolution of rights. Moving from culture to culture, property rights vary according to use, resources, and particular economic circumstances. Bailey and others have emphasized the benefits associated with the development of private property rights.\(^9\) Anderson and Swimmer (1997) broaden the discussion by emphasizing the importance of costs. They offer a more focused

\(^9\) See, for example, Anderson and Hill (1990), who study the American West; Alston et al. (1996), who study the Brazilian frontier; and Anderson and Lueck (1992), who study American Indian reservations.
original study that includes a cross-sectional empirical analysis of 40 early American Indian tribes. Their analysis supports the conclusion that rights of access to property vary depend on the relative value of the property resources and the different costs of establishing and enforcing property rights.\textsuperscript{10}

More modern studies of the evolution of property rights have focused on a number of closely related questions. First, can one explain which properties are owned publicly and which remain in the private sector; see, for example, Troesken (1997). Second, how important is the registration of land and the associated security of ownership important in the development of private property; see, for example, Miceli et al. (2001, Kenya), and Miceli and Sirms (2000, U.S.). Third, what affect did local government regulations of land use have on the rate of development of private property; see, for example, Fischel (1990, growth controls).

4.2.2. Creating individual rights in shared property

One important deviation from the classic “Coase Theorem” framework arises when resources are “common pool.” In such situations, access to a resource is at best only partly excludable; the resulting absence of entry controls leads to the “tragedy of the commons” in which the resource is overutilized. In essence, inefficiencies arise because there is no clear-cut mechanism by which individual agents can contract to exclude others and to prevent the dissipation or elimination of economic rents. Partial solutions to the tragedy of the commons involve means by which private property rights to common property are assigned. The empirical literature involves a series of qualitative studies, quantitative studies, and experiments that offer evidence as to the success of these private property regimes. Grafton et al. (2000) offer a relatively recent overview of the empirical literature, as well as a case study of a British Columbia fishery. Libecap (1989) describes the nature and complexities of the contracting issues that arise when one attempts to put into place property rights arrangements involving common resources.\textsuperscript{11}

One empirical approach has been to compare the success of various private property regimes on a cross-sectional basis within the U.S. Agnello and Donnelley (1975) find that oyster fisheries in states with private property regimes were more productive than those in states that had limited open-access regimes. Townsend (1990) and De Alessi (1998) both point to evidence that the regulation of common resource fisheries is not enough; it is essential to create appropriate property rights. Grafton et al. (2000), focusing on halibut fishing, conclude that it may take years for the efficiency gains from privatization to materialize, and that both short-run and long-run gains can be at risk if there are restrictions on private property rights (on the duration, transferability, and divisibility of those rights and by other inhibiting regulations). They also point to the

\textsuperscript{10} For further development of the theory underlie many of empirical studies, see, for example, Lueck (1994, 1995).

\textsuperscript{11} For an application to the study of oil fields, see Libecap and Wiggins (1984); for parking, see Epstein (2001).
potential adverse impact of preexisting regulations, such as rate-of-return regulations on public utilities.

The particular method used to create private rights to public resources can obviously make a difference. The use of transferable property rights has become quite popular in recent years, both with respect to the use of air resources (transferable emissions permits) and with respect to fisheries (transferable quotas). Transferable permits have generally been deemed successful. With respect to sulfur dioxide, each public utility was given an allowance for tons of emissions each year; the private right is transferable and bankable. As reported by Schmalensee et al. (1998), the result was an active market in emissions rights from the inception of the program in March 1997, and an improvement in economic efficiency.

4.3. Shared ownership

Institutions can facilitate exchange by defining and enforcing property rights and by enforcing contractual agreements. One institutional form which can in theory achieve commercial benefits is the cooperative, an arrangement in which individuals share ownership of common property. Such ownership can be enforced purely through contract, or through the exercise of coercive power by the state. A number of historical studies have evaluated particular institutional arrangements to better understand the specific factors that affect the costs of creating and enforcing property rights. Pirrong (1995) reviews a variety of commodity exchanges, concluding that as a general rule commodity exchanges were able to facilitate private arrangements to share ownership. Specifically, the commodity exchanges were able to achieve five goals: commodity measurement, contract enforcement, the policing of theft and fraud, and the mitigation of information asymmetries. The one exception was the Chicago Board of Trade, which failed to regularize its measurement of grains and to reduce significant information asymmetries between buyers and sellers. This empirical study, along with a number of others, has served to provide support for the view of Ellickson (1991) and North (1990) that private cooperation can facilitate private rights without the state playing an active role. Ellickson's study of cattle farming in Shasta County, for example, leads one to conclude that norms of cooperation that are based on mutual interests of affected parties can serve as the basis of a system of shared ownership, even if the parties do not interact repeatedly.

4.4. Property rights versus liability rules

The extent to which rights of property ownership are valuable to individuals depends in part on the mechanisms by which these rights are enforced. In a classic article, Calabresi and Melamed (1972) characterize the trade-offs between liability rules (enforcement

\[12\] With respect to air resources (sulfur dioxide), see, for example, Joskow and Schmalensee (1998) and Joskow et al. (1998); with respect to fisheries, see, for example, Grafton et al. (1996).
through suits for damage) and property rules (enforcement through injunctive remedies). Property rule regimes are generally seen as more efficient when transactions costs and the likelihood of strategic behavior are low, whereas liability rule regimes can be designed to minimize the possibility of strategic behavior and can work more effectively when transactions costs are high.\footnote{Calabresi and Melamed (1972) consider a third method for enforcing property rights—inalienability (rights that cannot be given away or sold, such as voting, human rights, and certain body organs). For a valuable discussion of the conditions under which alienability applies, see Rose-Ackerman (1985).}

A variety of empirical and experimental studies have been undertaken to evaluate the extent to which the law relies on one or both enforcement mechanisms and the effectiveness of those mechanisms. One set of applications has included empirical studies of the extent to which bargaining arises in property rights situations. For example, Farnsworth (1999) finds no evidence that parties bargain after judgments are reached in nuisance cases; this counters one's expectations that a judgment would serve to clarify the assignment of property rights. Apparently in the 20 nuisance cases under study the relationship among the parties was sufficiently acrimonious as to make bargaining difficult.

Another set of applications of considerable interest has been the normative and positive evaluation of situations under which the government has required the sale of private property. Thus, a regulatory taking can be seen as the imposition of a liability rule to enforce the ownership of certain types of private property, whereas property otherwise immune from a taking would be protected by a property rule.

While the takings debate has generated substantial normative debate, the empirical analyses of the pros and cons of government regulation of private property have been much more limited. A 1998 paper by Heller (1998) focuses on the “anti-commons problem.” In an anti-commons, multiple owners of a property each have the right to exclude others from the use of that resource, but no one has the privilege of sole use of the property. With too many owners exercising their right to exclude, the property is likely to be under- rather than over-used. Heller applies this theory to the utilization of property rights in Russia, explaining, for example, why many Russian storefronts were empty while street kiosks were full of goods. In an interesting empirical study, Miceli and Sirmans (2000) point out that even if it appears efficient to divide a property; that such a situation is likely to generate hold-out problems. They evaluate conditions under which the imposition of a “right to partition” the land, can (and cannot) improve social welfare. The authors also evaluate a remedy that requires the forced sale of the parcel of land, with the proceeds divided among the owners. They find that this remedy is often unfair and inefficient because it does not account for and preserve the subjective values that individuals may place on their particular parcels of land. They prefer a remedy in which the land is partitioned, and individual property ownership is maintained, despite the fact that there is a risk of strategic behavior and a potential cost of buying and reassembling the land.
In one intriguing study of the evolution of agricultural and fishing rights in Iceland, Eggertsson (1992) describes how initial private ownership of large tracts of land evolved into communal property. The resulting commons problem was responsible, in part (other causes included adverse trade relations, volcanic eruptions, pests and plagues, and a cooling climate) for the long-term economic decline of the Icelandic economy, running through the 18th century.

4.5. Consequences for efficiency and economic growth of poorly defined/enforced property rights

A system of poorly defined or inadequately enforced property rights can be expected, as a matter of theory to create economic inefficiencies and to inhibit or otherwise limit economic growth. In short, economic growth requires a government or governmental entities that establish property rights so individuals and firms can contract with a minimum of transactions costs.\textsuperscript{14} A variety of historical studies, covering a range of substantive areas, provide support for this proposition.\textsuperscript{15} With respect to water rights, the changes that occurred in the Western U.S. in the latter part of the 19th century offer a case in point. Kanazawa (1998) points out that California had adopted a system of water law that recognizes both riparian and appropriative rights. In riparian regimes rights are based on ownership of land and water rights are derived from use, not from physical possession. These rights are often less clear than appropriate rights, which are based on possession. Kanazawa offers empirical evidence that courts tended to promote appropriative claims when transactions costs were high to encourage the reallocation of water rights to their highest and best use.

Another focus in the property rights literature has been on the question of the extent to which property rights are essential if one is to improve the well-being of the poorest nations and their inhabitants. Using a cross-section sample of 47 countries with data focusing on changes from 1960 to 1989, Keefer and Knack (1997) find that deficient property rights institutions (as measured by indicators of the rules of law, the pervasiveness of corruption, and the risk of expropriation and contract repudiation) cause poor countries to fall back rather than catch up with their richer counterparts. More recently, using a cross-section of over 100 countries for the 1975–95 period, Norton (1998) finds that well-defined property rights improve the lot of both poorer countries and the poorest inhabitants of those countries.\textsuperscript{16}

In recent years, it has become more and more apparently that the specification and effective enforcement of a system of private property rights is essential if developing economies are to make a successful transition to a democratic capitalistic system which

\textsuperscript{14} The underlying theory is laid out, for example, in Norton (1998) and North (1981).
\textsuperscript{15} These studies include work by Barro (1991), Keefer and Knack (1997), and Knack and Keefer (1995).
\textsuperscript{16} Other recent studies with a geographical focus include Migon-Adbolla et al. (1991); Sub-Saharan Africa) and Lanjouw and Levy (2002; Ecuador). For a study of the importance of the effectiveness of government, see, for example, La Porta et al. (1999).
supports economic growth. For an interesting discussion of the difficulties of achieving a transition to a system of private property in Russia, see, for example, Shleifer, Boycko, and Vishny (1993, 1995).

5. The litigation process

5.1. Introduction: the costs of litigation

Litigation arises when the action of an injurer harms or allegedly harms a victim. The harm may be intentional as in the case of most crimes, accidental as in the case of many torts, or incidental as in the case of many nuisance situations. The behavior of injurers, and, in some cases, victims (their precaution and levels of activity) affect the frequency and extent of harm to one or both parties. Economic efficiency requires balancing the cost of harm against the cost of avoiding it, including the costs of risk bearing.

Most early studies of the costs of litigation were longitudinal in nature. In conducting those studies, social scientists focused primarily on the “supply” of social harm. In this framework, harms are seen not only as a function of the particularities of the legal system, but also as the consequence of more basic socioeconomic determinants. For example, automobile accident litigation may be affected by automobile registrations, gasoline prices, the average age of drivers, income levels, and urbanization. With respect to the legal system, differences among legal rules (e.g., strict liability vs. negligence), differences in rights of action, whether the trier-of-fact is a jury or a judge, and whether the attorney works on an hourly or contingent fee basis, are among the factors that can affect the costs of litigation.

Clark’s (1990) comparison of litigation trends in Europe and Latin America since 1945 is a useful example of a longitudinal study of litigation. Clark suggests different rates of litigation among Spanish regions from 1945 to 1967 or for Italian regions from 1952 to 1968 are best viewed as representing differential adjustments to varying environmental conditions. However, Clark and other similar longitudinal studies tend to avoid structural explanations that might reflect underlying sociological, economic, and political factors. Other than a passing reference to theories of the business cycle and to models of the evolution of courts, Clark does not look for variables that might characterize changes in the economy or to changes in the legal system that would produce more or less harmful behavior.

More recent time series studies of litigation have focused on the apparent “explosion” of litigation that occurred in the 1960s and 1970s. Analyses of these and more recent periods have typically been descriptive in style, emphasizing the substantive areas of law.

18 A conceptually more appealing analysis is that of Stookey (1990), whose model allows social crises to affect rates of dispute, which in turn affect litigation rates. Stookey notes that changes in litigation rates crate pressures for policy changes, which in turn induce changes in litigation rates.
in which litigation has grown most rapidly. For example, Galanter (1986) describes a rapid increase in products liability filings and in jury awards. Galanter also points to the fact that more than 98% of all civil cases are filing in state rather than federal court. As a result, any significant trends in litigation rates are likely to show up using state court data, which has been reported over the years by the National Center for State Courts. Finally, Galanter points to the interesting comparative fact that while many countries have lower overall rates of case filings than the U.S., that the per-capita use of the courts is in the same range in the U.S., Canada, Australia, England, Denmark, and Israel.  

5.2. The chronology of a legal suit

Initially, in the first stage of the chronology of a lawsuit, there is an underlying event in which one or more persons or entities (an "injurer" or "injuries") allegedly harms others (a "victim" or "victims"). The frequency of harm will generally be influenced by the choices that potential injurers make about the levels of activity in which they engage, and the precaution that they take when engaging in those activities. Activity levels are typically not easily monitored by public enforcers or by courts, so that much litigation (and the empirical evidence surrounding that litigation) centers on the alleged injurer's level of precaution. The activity levels and extent of precaution chosen by those who are potential victims may also affect the frequency of harm.

As will be true at each stage of the litigation "game," the "rational" decisions of potential injurers and victims will be dependent on their expectations as to the stream of benefits and costs associated with their activity and precaution choices. These benefit and cost streams, are of course, related directly to the decisions that are likely to be made in each of the further stages of the litigation game.

In the second stage of a legal dispute, the party that allegedly suffered harm typically hires an attorney and then decides whether and where to assert a claim. If made on a rational and self-interested basis, these decisions will be the solution to a sequential game that balances the costs of asserting the claim against the benefits that might be expected through settlement or trial. The extent to which the attorney's interests are or are not aligned with those of the client is a significant issue for any legal system. Moreover, this principal-agent problem is likely to be significantly influenced by the fee arrangement under which the attorney operates.

The third stage of litigation occurs after the claim has been filed, but before trial. Pre-trial litigation efforts, which relate generally to the discovery process, include responding to complaints, answering interrogatories, and taking depositions. By providing information, the discovery process serves to help the parties sharpen their expectations about trial outcomes. Consequently, the pre-trial discovery process provides the central

---


20 For a general overview of the litigation process, with an emphasis on normative issues, see Cooter and Rubinfeld (1989); see also Kaplow and Shavell (2002).
input for the fourth stage of the litigation process, settlement bargaining. If the parties can reach a cooperative outcome is reached the case settles; the non-cooperative alternative is a trial.

The fifth stage of the litigation process is the trial itself. Trials can be costly exercises, although costs are likely to vary widely depending on the stakes in the case, and the nature of the forum in which the trial takes place (jury vs. judge trial, arbitration vs. litigation, etc.). Because litigation is costly, trials are typically negative sum games, with a possible exception that can arise when external benefits to one party make the payoffs of the trial game substantially asymmetric.

It is useful for our purposes to also distinguish a sixth stage of the litigation game—the trial outcome itself. Judgments by judges or juries can be injunctive or compensatory or a combination of the two. If both parties are satisfied with the judgment, the litigation process ends. However, one or both parties can opt for the seventh stage—an appeal to a higher court or courts.

The theoretical game-theoretic literature describing each of the stages of the litigation process, and/or combinations of several stages is substantial. However, the empirical evidence that is needed to provide the foundation for those models is in much more limited supply. In the subsections that follow, we highlight some of the relevant empirical literature. Because our focus is on the litigation process itself, we begin with the filing of the lawsuit.

5.3. Forum and law choice

An individual who believes that he or she has suffered harm must (with the assistance of an attorney) choose the appropriate law and the appropriate forum in which to file. It is not unusual for the victim to have a range of law and forum options, depending on whether state, federal, or foreign statutes apply. Moreover, defendants face forum choice options as well. Thus, a defendant facing a patient infringement claim in a jurisdiction believed to be sympathetic to the plaintiff may choose to file a declaratory judgment claim in a more friendly location.

The ability and the incentive to forum shop are both likely to be influenced by the decision rules that courts use to resolve conflicting forum claims. There are two related questions of interest. First, suppose that a claim is filed in one jurisdiction involving parties that operate in several jurisdictions. Which jurisdiction’s law should apply? Second, suppose that the parties file related claims in two distinct jurisdictions. How do the jurisdictions decide which should take precedence?

The former question has been the focus of much legal debate and some empirical analysis. Indeed, the legal footing on which the choice of law (or alternatively the conflict of laws) doctrine relies has evolved over the past half century. Under the First Restatement of Conflict of Laws (1934), for example, solutions were likely to depend on non-legal factors, rather than on the substantive laws of the conflicting jurisdictions. The Second Restatement (1971) gives credence to the laws of the possible forums, but suggests taking into account a wide range of factors, including, but not limited to: (i) the
relevant policies of each forum; (ii) the protection of justified expectations; (iii) the certainty, predictability, and uniformity of results; and (iv) the ease of determination and application of the law.21

Under the Second Restatement, judges are afforded substantial discretion in making choice of law decisions. In a study of 802 multi-court opinions involving choice of laws, Borchers (1992) found a wide variation in the pattern of results. Courts that tended to rely on the First Restatement tended to favor rules that favored plaintiffs’ recovery less often than those that relied on the Second Restatement. In addition, courts that relied on the First Restatement were less likely to benefit local parties than those that did not. On the whole, Borchers reaches the conclusion that whatever the cited basis, most courts use their own judgment, in the process finding a way to do what they believe is substantively appropriate without necessarily following the law. However, Thiel (2000) comes to a contrary conclusion, based on a reanalysis of Borchers’ data using more sophisticated econometric techniques.22 Thiel finds little evidence to support the theory that judges are more favorably inclined to residents or to plaintiffs, although he does (not surprisingly) find that judges tend to favor the law of their own forum.

Thiel’s article is (to our knowledge) the first that makes an effort to draw empirical implications from theories of interest-group politics. He suggests that the competition among jurisdictional laws can be seen as dynamic prisoners’ dilemma game, in which the cooperative outcome (which maximizes the joint welfare of the two jurisdictions) may or may not be achieved.23

Whether forum shopping enhances or diminishes economic efficiency is open to debate. The positive case for shopping is based on the possibility that a party can choose a forum that will allow from a quicker and more accurate trial. For example, a number of authors have argued that Delaware’s successful effort to develop a specialization as a center for corporate law has generated a positive “race-to-the-top,” not a race-to-the bottom.24 The case against forum shopping is based on the view that differences in expected awards across forums reflect true differences in the cost of failing to take precaution. Suppose, for example, that the cost of precaution is sufficiently high in the jurisdiction in which a plaintiff is injured so as to make the injurer non-negligent in tort. If the plaintiff were to file in a jurisdiction in which precaution is less costly, the defendant will be overdeferred and litigation costs will be excessive.25

There is evidence that forum shopping has increased substantially over the past thirty years. For example, the fraction of state court cases that were removed to federal court

22 Thiel uses a two-limit Tobit model, which builds on the assumptions that (1) some choice of law arguments are not made by each party to the litigation because the interest is not sufficient to cross an appropriate threshold, and (2) judges can only apply choice of law analyzes when the issue is raised by one or both parties.
23 For another study that takes an interest-group, political science perspective, see Solimine (1989).
24 See, for example, Bebchuk et al. (2002) and Romano (2002).
25 For a normative discussion of forum shopping along with empirical evidence, see Clermont and Eisenberg (1995).
increased from 15% in 1970 to over 30% by the year 2000. This is consistent with the view that defendants often choose more friendly forums after plaintiffs have filed suit. The alternative explanation is that defendants are simply responding to an inefficient forum choice by plaintiff. Clermont and Eisenberg recently attempted to sort out a number of competing interpretations of the increased forum shopping. They studied a large sample of cases (utilizing a database obtained from the Administrative Office of the Courts of three million federal cases terminated over thirteen years) involving diversity of jurisdiction; in such cases federal as well as state court venues are available to the parties. While plaintiffs were successful in 71% of cases overall, their success in cases removed to federal court was only 34%. This was due in part to the ability of defendants to move cases away from plaintiff-friendly jurisdictions (which may or may not be efficiency-enhancing). However, it also reflects a case selection effect, since to some extent the cases that were removed were relatively weak from plaintiff’s perspective, wherein the strong plaintiffs’ cases had been settled or otherwise resolved. Using regression analysis to control for characteristics of cases and hopefully the case-selection effect, the authors concluded that forum-shopping effects remain substantial. Forum selection through removal, all else the same, reduces a plaintiff’s probability of success from 50% to 39%.

Because of their focus on diversity cases, Clermont and Eisenberg make the plausible presumption that the dominant advantage of forum shopping inures to the defendant. Their view is that the transfer of forum causes some plaintiffs to abandon their cases or to settle on relatively unfavorable terms; moreover, it also makes litigation more difficult for the plaintiff. Yet, absent the ability to transfer, forum shopping clearly can and has benefited plaintiffs. One recent study lays out the substantial benefits to both parties from forum shopping in patent litigation cases.

In summary, empirical evidence supports the view that there is substantial forum shopping by plaintiffs, whose effects are diminished but not eliminated by defendants’ ability to transfer cases. These effects appear significant, even after controlling for important case and selection effects. As with other areas of litigation, there remain many unanswered questions related to the choice of laws and forum. To what extent do defendants make an initial strategic forum choice, as, for example, by filing for declaratory judgment? What are transactions costs that are imposed by the forum-shopping process? What risks are put on the parties by the uncertainty surrounding choice of law and choice of forum? While the current state of empirical research is useful, more will be needed before the normative issues relating to forum and law choice can be resolved.

---

27 This discussion is based on Clermont and Eisenberg (1998), which reports results from their 1995 paper.
28 The authors also examined cases that were transferred from court to another, rather than removed from state to federal court. Plaintiffs’ probability of winning in transferred cases, other things equal, falls from 50% to 40%.
5.4. Pre-trial

The litigation process is best viewed as a non-zero sum game in which each of the parties reveals information about the strength of its strategic threat point. It is widely accepted that trials tend to arise when the judgment expected by the plaintiff is greater than the judgment expected by the defendant, i.e., when the plaintiff is relatively optimistic. Indeed, in order for settlement to be possible, the savings in costs to the parties must exceed their relative optimism about trial.\(^{30}\) Thus, settlements tend to occur when relative optimism is small, whereas trials tend to occur when the plaintiff expects a large trial judgment and the defendant expects a relatively small judgment. Typically, one would expect the pre-trial discovery process to move the discovering party’s expectation closer to the trial outcome that would occur if all information were pooled, while at the same time reducing the variance of the distribution of expected values, thereby reducing the risks that the litigating parties face.

Any pre-trial information that decreases the mean of the plaintiff’s distribution of possible trial outcomes will make the plaintiff less optimistic and thereby decrease the probability of trial, and conversely for the defendant. Moreover, because trial occurs in a relatively small number of cases where the plaintiff is optimistic relative to the defendant, a reduction in variance caused by the pooling of information is likely to result in fewer trials as well. It is useful to view the pre-trial discovery process as including the voluntary disclosure of information, which will tend to support each side’s optimism, and the mandatory disclosure, which will have the opposite tendency.\(^{31}\)

Pre-trial discovery can satisfy an additional goal, by increasing the likely accuracy of any settlement that the parties reach. An accurate settlement is a hypothetical settlement that would be reached if the parties had complete information about the law and the facts of the case, and the defendant paid an amount equal to the complete information judgment to the plaintiff (or a correspondingly appropriate injunction were imposed if an equitable remedy were required). Similarly, discovery can increase the accuracy of trials by improving the information available to the court.

Finally, the legal rules surrounding pre-trial processes also have implications for the incentive of the parties to take precaution and on the probability that harm will occur in the first place. Discovery that is likely to increase the relative optimism of the plaintiff (or the relative pessimism of the defendant) is likely to increased the expected value of a claim, and thereby the likelihood that a case will be filed.\(^{32}\) This will, in turn, increase deterrence by increasing the potential defendant’s incentive to take precaution. Discovery could encourage plaintiffs to drop claims that have little merit, i.e., claims that were

---

30 For an overview of the economics of discovery, see Rubinfeld (1998) on which portions of this subsection are based; for a more formal analysis, see Cooter and Rubinfeld (1994).
31 For a general discussion of voluntary vs. mandatory disclosure see Hay (1994) and Shavell (1989).
32 This abstracts from changes in the probability of settlement; if plaintiff becomes more optimistic, the probability of trial could increase, which, other things the same, could diminish the expected value of the claim (because trial costs could be large).
based on false optimism. In general, broad discovery rules are likely to increase deter-
rence to the extent that they improve the specificity of the litigation process, which has
the effect of increasing the gap in liability costs between failing to take care and taking
care.33

The theory of pre-trial discovery makes clear the trade off between the costs of pre-
trial discovery and the costs of trial. A more extensive and more costly discovery process
is likely to increase the probability of settlement (which is typically well over 90%) and
thus to save on trial costs. However, discovery can itself be highly costly, and the ap-
propriate balance is not clear. If discovery had no effect on settlement, the cost of obtaining
information through discovery should be around 10% of the cost of finding that infor-
mation through trial in order for discovery to save transactions costs directly. However,
discovery can increase the likely of settlement, which itself reduces transactions costs,
so the ultimate trade-off is less transparent. Moreover, discovery can be used strategi-
cally by the parties; since discovery costs are typically borne by the responding party, it
can be advantage for a party requesting discovery to engage in broad discovery simply
to raise its rival’s costs.

Interestingly, there is substantial variation in discovery practices across different le-
gal environments, which offers a potential and relatively untapped source of empirical
information that is relevant to the normative questions that have just been asked. We
know that discovery practices (and costs) vary with legal environments, based among
other things, on (i) whether the system is adversarial as in the U.S. (a common law
country), or inquisitional, as in most European countries (which operate under civil
law); (ii) whether there is substantial reliance on juries as triers-of-fact, as in the U.S.,
but not England34; (iii) whether there is substantial fee shifting from losing to winning
parties; as in the U.K. and parts of Europe, but not in the U.S.; and (iv) whether the un-
derlying culture relies heavily on attorneys, and relies on the court system to encourage
the provision of information and to resolve disputes, as in the U.S., but not in Japan.35

Some historical, primarily non-econometric evidence supports the view that discov-
ery efforts are responsive to economic incentives. Thus, there is a strong finding of a
positive correlation between the stakes of the case and discovery effort.36 Surprisingly,
however, one study found that cases with greater discovery tended to settle less than
those with less discovery.37 While the author did control for case complexity and case
size, the formal analysis did not reflect the relatively complex set of strategic incentives
facing the parties.

33 P’ng (1987).
34 There is some empirical evidence that delays in getting to trial are on average greater for judge trials than
for jury trials (Eisenberg and Clermont, 1996), and alternative dispute resolution processes do not reduce
delay (Heise, 2000).
35 Langbein (1985) offers an interesting qualitative comparison of discovery in the U.S. and in Germany. For
a comparison with Japan, see Ramseger and Rasmussen (1997).
36 See, for example, Trubek et al. (1983), pp. 95–96 and Glaser (1968). For a more general overview of the
non-econometric empirical evidence, see McKenna and Wiggins (1998).
37 Glazer (1968).
On the whole, there has been very little econometric work on the pre-trial behavior of the parties. One important exception is Shepherd (1999). In a study based on an interview survey of attorneys in 369 federal civil cases, Shepherd estimated a simultaneous-equations tobit model that emphasized the interdependent nature of the strategic discovery choices made by plaintiffs and defendants. He modeled each party's discovery effort as a function of the expected recovery in the case, the expected non-monetary recovery, various case-specific factors, the fee-arrangement with the client, and the discovery effort by the opposing party. Shepherd found that plaintiffs typically choose their discovery independently of the defendant, i.e., they do not respond strategically. Defendants, however, respond almost entirely to the plaintiffs' discovery efforts in a manner consistent with a tit-for-tat game-theoretic strategy. Finally, plaintiff attorneys operating under hourly fee arrangements tend to conduct on average more than 5 days more discovery than do attorneys operating under a contingency. Whether these and other results will hold up in future studies that use data sets that allow one control more effectively for case characteristics and to measure discovery effort objectively remains an open question.

To our knowledge, there has been little or no research into the costs imposed on the larger judicial system by the discovery process. Thus, we know little about the costs to the parties that could associated with the revelation of confidential information, or the costs to non-litigants who might be affected by case outcomes. Nor, do we have much information on "discovery abuse," which might be defined as discovery requests where the direct information benefit to the requesting party is less than the cost to both parties of meeting that request.

Much of the empirical analysis of pre-trial discovery has focused on the benefits and costs of changes in discovery rules, with the most recent evidence flowing from a debate as to whether recent changes in federal discovery rules that required substantial disclosure by both parties.38 Because half of the local districts opting out of the required changes, as allowed under the reforms, a more or less natural experiment (but-for selection effects) was created. Key questions were (i) would mandatory disclosure increase or decreasing the costs of litigation, and, (ii) would mandatory disclosure delay or speed up the litigation process?

One recent study by the Federal Judicial Center surveyed 2,000 attorneys involved in 1,000 cases that were terminated in 1996.39 The survey suggested that attorneys, on balance, believed that the accuracy of the process was improved, which would tend to lower costs. Moreover, there was evidence of a small decrease in the time involved in litigating cases. A second study, conducted by the RAND Corporation, compared the behavior of a small group of district courts prior to the discovery reforms, some of which

38 The majority of studies of the role of discovery have been undertaken primarily by socio-legal scholars. See, for example, a Federal Judicial Center study (Connolly et al., 1978), the work that emanated from the Wisconsin Civil Litigation Research Project (Trubek et al., 1983), or more recently a study by the National Center for State Courts (Kelliz et al., 1993).
39 Williging et al. (1997).
had a form of mandatory disclosure and some of which did not.\textsuperscript{40} In this case, the RAND group found no significant differences in delay or litigation costs, but some increased dissatisfaction of attorneys who operated under mandatory disclosure rules. Whether changes in disclosure rules really have little effect, or whether we simply have not been able to measure those effects with sufficient precision, remains an open question.\textsuperscript{41}

5.5. Settlement

We suggested previously that cases are less likely to settle the greater the relative optimism of the plaintiff about the trial outcome. More generally, the basic economic model of litigation suggests that settlements will occur if there is a positive cooperative surplus to be enjoyed by settling.\textsuperscript{42} This surplus consists of three elements (1) the sum of the trial costs of the two parties; (2) the difference between the plaintiff’s and defendant’s subjective expected payoffs from going to trial; and (3) the sum of the risk-bearing costs borne by both parties. To put this formally, let

\[ T_p(T_d) = \text{subjective value to plaintiff (defendant) associated with a successful trial outcome}; \]
\[ c_{tp}(c_{td}) = \text{cost to plaintiff if the case is settled.} \]

Then, according to theory, cases will settle when \( G = (T_p - T_d) + (c_{tp} + c_{td}) > 0 \).

This model of settlement rules out the possibility of strategic behavior by one or both parties; such behavior must be the result of an effort by the parties to increase their share of the cooperative surplus from settling the case. One practical approach to account for strategic behavior is to add randomness to the settlement-trial decision, in which case the probability of settlement is a function (usually a cumulative distribution function) of the size of the cooperative surplus, measured directly or proxied by a series of case-specific variables. A substantial number of empirical studies of the settlement-trial decision have been authored over the years. The earliest studies by socio-legal scholars have been both longitudinal and cross-sectional. Only in the past decade or two, however, have economists begun the task of specifying and estimating structural models of the behavior of the parties, and they have done so to answer a variety of interesting questions.

One question is whether augmenting compensatory damages will result in more or fewer trials. Theory tells us to expected three competing effects: (1) the tendency of relative optimism to cause cases to go to trial will be strengthened, which will reduce the probability of settlement; (2) the cost of trying cases will increase, which will increase the probability of settlement; and (3) the risks faced by both parties will increase, which

\textsuperscript{40} Kakalik et al. (1998).
\textsuperscript{41} For another study that found no effects, see Huang (2000).
\textsuperscript{42} For overall reviews of the relevant theoretical literature, see Cooter and Rubinfeld (1989), Daugherty (2000), Hay and Spier (1998), and Kaplow and Shavell (2002).
should increase the probability of settlement. In one study of a large number of antitrust trials, Perloff and Rubinfeld (1987) found evidence suggesting that, where reputation effects are important and where the parties tend to be pessimistic, treble damages leads to a decrease in the proportion of settled cases and an increase in the number of trials.

Other studies, however, have come to contrary conclusions, including Danzon and Lillard (1983), who studied medical malpractice claims. Danzon and Lillard model four equations: two trial equations explain the probability of plaintiff winning and the amount of the verdict, while the settlement equations explain the minimum demand of the plaintiff and the maximum offer of the defendant. The authors assume that increased stakes increase the randomness in the model proportionally, but the cost of litigation less than proportionately. This leads to their conclusion that higher stakes will cause more cases to be litigated. They also find the higher the plaintiff’s probability of winning at trial the lower the probability that the case will be tried.

Recent studies of settlement behavior have become more sophisticated in their ability to account for selection effects and the close link between settlement behavior and trial outcomes. Viscusi (1988) focuses on the decisions to drop and to settle product liability claims. He concludes that the expected award had from two to nine times the influence of the variance on settlements. Fournier and Zuehlke (1989) estimate a plaintiff’s settlement demand equation that explicitly accounts for self-selection, using a survey of civil filings from 1979–81. They find that higher trial awards and increased variance both increase the probability of settlement.

Viscusi and Scharf (1996) account for selection effects at both the drop and the settlement stages of litigation using a large number of product liability cases. They find that higher trial awards have a greater (positive) effect on the defendant’s settlement offer than on the plaintiff’s request. Conversely, higher probabilities of winning at trial have a greater effect on the plaintiff’s request than the defendant’s offer. The first result is consistent with a defendant that is more risk averse than a plaintiff, whereas the second result suggests the converse.

Using the same data set described previously, Perloff, Rubinfeld, and Ruud (1996) estimated a joint model of trial outcomes and settlements. They modeled settlements as a function of the expected distribution of outcomes, conditional on a case going to trial. In their model the probability of settlement is a function of the probability that plaintiff and defendant expect to win at trial as well as the variance of the trial outcome. They find that risk aversion plays an important role in explaining why cases settle instead of going to trial. When the probability of winning is near 50%, the variance of trial outcomes is at its maximum. When the probability of success increases above 50%, the variance decreases. Using this fact, and accounting for the importance of risk aversion, the authors find that for every 1% increase in the probability that the plaintiff wins, the probability that the case settles increases by nearly 0.13%. Moreover, because

\[ \text{In a binomial distribution with success probability } p, \text{ the variance is } p(1 - p), \text{ which is maximized at } p = 1/2. \]
the magnitude of the risk aversion effect increase with the size of damages awarded, trebling antitrust damages has a dramatic effect on the probability of settlement. Once again, were treble damages eliminated, the fraction of cases going to trial would increase substantially.

While reputation effects per se may or may not be significant in particular cases, the likelihood that settlement will be reached may be a function of the nature of the interactions among parties. In one recent study Johnston and Waldfogel (2002) found even when litigants have no history, that attorneys may be involved in repeated play. Using a data set on 2,000 federal civil cases filed in the Eastern District of Pennsylvania in 1994, they found that a history of attorney repeat interaction has a significant positive effect on the probability that cases settle and that such interactions tend to shorten the disposition of the litigation generally.

The particulars of settlement behavior depend on the nature of the parties (firms or individuals, risk neutral or risk-averse, etc.), on the nature of the cases (large stakes, small stakes, reputation effects, etc.), and more generally on the institutional characteristics associated with the particular subject matter at issue. We have cited previously a range of studies of civil litigation generally, and of torts, antitrust, products liability and medical malpractice. However, the range of cases covered almost all areas of civil litigation, including tax and securities.

Settlement is often seen as the preferred outcome from a policy perspective because it avoids trial costs. However, seen from the perspective of the litigants themselves, the preference for settlement is not so clear. On one hand, Gross and Syverud (1996) explain, relying in part on a survey of attorneys, the pervasive preference for settlement in most legal systems. On the other hand, Galanter and Cahill (1994) point out that efforts by the courts to encourage settlement, such as court-ordered mediation, are met with mixed satisfaction by the litigants.

5.6. Trial

The outcome at trial is the result of a complex interaction between the efforts that both parties put into the trial and the underlying facts and law of the case. Perhaps the most central topic of empirical analysis has been the fraction of cases that are won by plaintiffs and by defendants at trial. All researchers working on this question are aware that

---

44 Other relevant medical malpractice studies include Hughes (1989), Sloan and Hoerger (1991), and Farber and White (1991).
45 Lederman (1999) studied 400 Tax Court cases, and found significant selection effects. Alexander (1991) found that fewer than five percent of litigated securities cases were tried to judgment.
46 See also Gross and Syverud (1991).
47 Institutional constraints can affect settlement rates as well. Based on an analysis of the timing of settlements of automobile bodily injury insurance claims, Kessler (1996) found that delay in the trial court system generally increases delay in settlements, and that tort reforms designed to reduce delay in settlement did not work as intended. Kessler used two separate cross-sections of settled claims, one from 1977 and one from 1987.
the concepts of winning and losing are not well defined, since both may be relative rather than absolute concepts. The plaintiff "wins" a civil suit, in one sense, if the court awards damages or provides injunctive relief. Many civil suits, however, concern not the fact of defendant's liability, but its extent. From this perspective, the plaintiff "wins" at trial only if the damage award is larger than the defendant's settlement offer.

In any case, it is well known that the selection of cases for trial need not, indeed, is unlikely to be, a random selection of filed cases. Despite the former, and in light of the latter, a number of papers have focused on whether and under what conditions one should expect plaintiffs to have a 50% win rate. Priest and Klein (1984) have received prominent treatment in this debate, in part because they were careful to think through the selection-bias issues that arise when one accounts for the fact that those cases that go to trial are likely to be those for which there is no (positive) cooperative surplus. Their starting point is that one should expect a 50% win rate; if the rate tended to be higher, for example, defendants would find it in their advantage to settle weaker cases, and vice versa.

Others, however, have pointed out that the interests of the parties and their attitudes towards risk may not be symmetric, which could generate win rates that are significantly different from 50% for plaintiff. For example, if one or both parties are concerned about reputation effects, the benefits of winning may be greater than those that are trial specific. Arguing that defendants are likely to be more concerned with reputation than plaintiffs, Perloff and Rubinfeld (1987) found that approximately 70% of all antitrust cases in their data set were won by defendants. Wittman (1985, 1988) argued further than when the parties disagree about the expected trial award, win rates can deviate substantially from 50%. Perhaps the general points here are best summarized by Shavell (1996), who points out it is possible for the cases that go to trial to result in plaintiff victory with any probability, and consequently, that the probability of plaintiff victory in settled cases, had they been tried, may be any other probability.

Waldfogel (1995) has offered a thorough and broad empirical analysis of the selection of cases for trial.48 Using data describing contracts, intellectual property, and torts cases filed in the Southern District of New York from 1984–87 and closed by 1989, Waldfogel finds strong statistical relationships between the rate at which cases go to trial and plaintiff win rates. When relatively few cases go to trial, the plaintiff win rate tends to be close to 50% as predicted by the model of selection bias. However, when the trial rate increases, which is presumably due to significant asymmetries between plaintiff and defendant expectations about trial outcomes, the plaintiff's success rates diverging from 50%, with rates being higher in some cases and lower in others. Because of data limitations, Waldfogel is unable to explain the underlying source of variation in win rates across case types.49

48 For an earlier analysis of similar cases, see Eisenberg (1990).
49 See also Siegelman and Donohue (1995), who discuss the complexities of sorting out settlement decisions and trial outcomes in the context of employment disputes.
In a somewhat more recent paper, Waldfogel (1998) uses the same data set to evaluate the divergent expectations explanation of why cases go to trial (as discussed previously) with an alternative asymmetric information theory. Under the latter theory one party, perhaps the defendant, knows the probability that plaintiff will win at trial, while plaintiff knows only the distribution of plaintiff victory probabilities. If the uninformed plaintiff makes a settlement demand, the defendant will tend to accept that demand only if the plaintiff’s probability of success is relatively high. As a result, for those cases proceeding to trial, the plaintiff’s win rate should be consistently below the fraction of plaintiff winners in the pool of all cases, and presumably smaller than 50%. Waldfogel’s evidence rejects the asymmetric information theory in favor of the divergent expectations theory discussed previously.

There is little doubt, as Eisenberg and Farber (1997) point out, that the process by which lawsuits get filed and go to trial depends not only on the subjective expected value of the claim, but also psychological and other non-pecuniary motives. In a broad sense, those cases that are filed and that eventually go to trial, are likely, other things equal, to be those where the costs of litigation are lowest. Interpreting costs broadly, we can say that a “litigious” plaintiff is one that views the costs of going to trial to be unusually low. Using a data set on over 200,000 federal civil litigations, the authors find that plaintiffs who are individuals are likely to have higher rates of trial than corporations (lower settlement rates), and consistent with the implied asymmetry of interests, lower plaintiff win rates.50

The costs of litigation are in part a positive function of the time it takes to dispose of a case at trial or through settlement. A number of economists and socio-legal scholars have studied the factors that affect the time of disposition of civil trials. One important effort along these lines has been made by researchers at the RAND Corporation’s Institute for Civil Justice. In evaluating the Civil Justice Reform Act of 1990, RAND studied one year of civil jury case outcomes from 45 of the 75 most populous counties in the U.S. The study found that the average disposition time for a case that went to trial in Cook County Illinois, was over five years.51

A useful overview of trends in civil jury verdicts for the decade beginning in 1985 is offered by the Moller (1996) RAND study. Focusing on all civil jury verdicts reached in the state courts of general jurisdiction in 15 jurisdictions, Moller found, among other things, that plaintiffs win slightly more than half their cases. However, the win rate ranged from a high of 66% in automobile personal injury and business cases to lows of 44% in products liability cases, and 33% in medical malpractice cases.52

Case outcomes often vary from jurisdiction to jurisdiction, and from data set to data set. Using a national data base on civil filings obtained by the Administrative Office of the Courts, Clermont and Eisenberg (1992) found win rates of close to 50% for products

50 For a discussion of the possible influence of judicial ideology on case outcomes, see Ashenfelter et al. (1995).
52 See also Peterson and Priest (1982), Shanley and Peterson (1983), and Hubbard (1987).
liability and medical malpractice cases. Surprisingly, however, they found win rates to be substantially higher before judges than before juries, despite the fact that plaintiffs select judge trials in only about 10% of the cases. Whether all of this can be explained by selection effects arising when cases vary substantially on many dimensions, remains an open question.

There is a long-line of literature evaluating the relatively abilities of juries and judges to perform their evaluative functions. Vidmar (1998) provides a useful overview of the jury literature. He notes that in 1993 that only 1.8% of all civil cases filed in U.S. District Courts involved jury trial, and the fraction is very similar (2.0%) for state cases. He reviews and evaluates issues such as the ability of the jury to evaluate expert testimony, to comprehend complex legal arguments, and to understand jury instructions. With respect to the question of whether juries have a pro-plaintiff bias, the literature reaches a negative conclusion.

A recent subject of intensive study and debate has been the ability of juries to evaluate claims for punitive damages. It is well known that the distribution of such awards has a very high variance. What is less settled is the question of whether those awards are related systematically to factors raised by the issues in the case, or whether they are largely due to the particularities of juries. In one study, Hastie et al. (1998) used responses to hypothetical questions about the appropriateness of punitive damages by experimentally-chosen jurors. Vidmar criticizes this study for, among other things, focusing on the wrong questions, but is somewhat more sympathetic to Hastie and Viscusi (1998), who conclude in from a related study that juries are less able than judges to manage punitive damage deliberations.

The influence or lack of influence of judges on trial outcomes has itself been a subject of some intensive empirical work. In evaluating more than 200 district court cases involving the constitutionality of the U.S. Sentencing Commission in 1988, Cohen (1991) concludes that at the margin judges are likely to promote their own self-interest. Using data extracted from Jury Verdict Research's Personal Injury Verdicts and Settlements (59,304 trials and 27,429 settled cases), Helland and Tabarrok (2000) report than much of the difference in win rates (higher with judges) and awards (higher with juries) between juries and judges can be explained away by case selection effects.

5.7. Appeal

Relatively little empirical work has concentrated on the appellate process. One important exception is Kessler et al. (1996). The authors take on the question of whether and

---

53 Important if only for its historical significance is the class study of juries and judges by Kalven and Zeisel (1971).
54 See also Viscusi (2001).
55 For one interesting study relating to the D.C. Circuit and environmental appeals, see Revesz (2001).
56 In a related paper, Tabarrok and Helland (1999) show that plaintiffs are more likely to win tort awards and that the awards are likely to be higher in states in which state judges are elected in partisan races than in other states (non-partisan elected judges or appointed judges).
to what extent one should expect deviations from a 50% win rate with respect to appeals, emphasizing that there are a range of factors that empirically explain deviations from the 50% result. Analyzing 3,529 appeals decided by the Seventh Circuit Court of Appeals between 1982 and 1987, they find several case characteristics that influence win rates in the manner than economics would predict. They include: differential stakes (the party that has the most at stake is more likely to win), differential information (the better informed the party, the higher the win rate), legal standard favoring one side (obvious), and agency effects (the party represented by an hourly fee lawyer will have a lower win rate at trial because the attorney will have a higher probability of taking the case to trial).

Building on our earlier discussion of differential stakes, there are reasons to expect win rates substantially below 50%. It seems clear that the probability of success on appeal is likely to be quite low for the entire population of tried cases. If the stakes (and risk aversion) were the same for both parties, then one would predict a 50% success rate on appeal. However, it seems reasonable to expect that the party that pays the cost of an appeal does so because the potential reward is quite high and the cost of an appeal is low (at least in relation to the cost of the trial itself). The implication is, given sample selection, a win rate of substantially less than 50%. Indeed, that is what Clermont and Eisenberg (2002) find in a recent analysis using the Administrative Office of the Courts database described previously. Defendants are successful in reversing loss jury trial decisions in 31% of their appeals, while losing plaintiffs success in only 13% of their appeals from jury trials. The authors point out that these results cannot be fully explained by sample selection, appealing instead to the argument that appellate judge’s tend to be more favorably disposed to defendants than are either trial judges or juries.57

5.8. Alternatives to litigation: arbitration and mediation

Throughout the years, a wide variety of litigation reforms have been proposed and in some cases promulgated to streamline the litigation process. We have previously described a number of studies of the relatively recent attempts at the reform of the discovery process. In this subsection, we briefly highlight some of the empirical work that has focused on alternative dispute resolution systems, including arbitration and mediation.

Many studies of both arbitration and mediation have focused on the effect of various interventions on the perceived fairness of the process, rather than the economic effects on settlement behavior. Thus, in one recent study of court-connected civil case mediation programs in nine Ohio state courts, Wissler (2002) did not attempt to look at whether settlement rates themselves increased as a result of the mediation (45% of mediated cases did settle).58 She did, however, find that few cases went to trial whether

57 The differences are less substantial with respect to appeals of decisions by judges (23% success for defendants and 18% success for plaintiffs).
there was mediation or not—3% of mediated cases and 2% of non-mediated cases were actually tried. In addition, 2% of the mediated cases and 4% of the non-mediated cases were resolved by arbitration. 59 Interestingly, she also found that when mediators recommended settlement, cases were more likely to settle, while at the same time the parties were more likely to view the process as unfair. In sum, one cannot reject the possibility that successful mediated cases were those in which the parties’ expectations were least divergent; it follows that successful mediated cases were those most likely to settle without mediation.

Another study focused on a four year study in the Northern District of California of an alternative dispute resolution process called early neutral evaluation. In the “experiment” half of the cases of certain types were assigned to a mediator and half were not. Rosenberg and Folberg (1994) report substantial satisfaction with the process. Moreover, in about half the early neutral evaluation cases parties reporting savings in litigation costs and shorter disposition times.

The mediation studies just described all involved interventions by the courts. An alternative form of dispute resolution is private in nature, with the parties agreeing to pay a fee to a third-party to assist with the resolution of the dispute. A useful study of these forms of mediation was offered by Rolph et al. (1994). The authors studied a wide variety of civil cases that were active in Los Angeles, California in 1992 and 1993, relying on interviews with parties and mediators, as well as lawyers’ case records and Court records. They find that arbitration rather than mediation is the most popular form of alternative dispute resolution ("ADR"). While they find little or no overall effect on the civil caseload of the Los Angeles courts, they do express optimism that private forms of ADR can be successful in the future (91 former judges were offered private dispute resolution services at the time of the study). 60

With respect to arbitration, it is typically thought that such programs can reduce delay and costs by providing a more efficient substitute for trial. But, as MacCoun (1991) has pointed out, because most disputes are resolved without trial, such programs are likely to divert more cases from settlement than from trial. The implication is that arbitration (and for that matter mediation as well) may actually increase delay and congestion in the courts. MacCoun found such a result in a study of court-annexed automobile arbitration in New Jersey. After the introduction of the arbitration program there was no reduction in the rate of cases going to trial, and an increase in the time of disposition of cases. He concludes for arbitration, as others do for mediation, that alternative dispute resolution programs appear to offer improvements in the fairness of the process, but do not necessarily increase economic efficiency, and in some cases, may make it worse. 61

There are clearly some who believe that neither private nor public interventions of the type just described are likely to much success if any. Whether we should stick with

59 p. 669.
60 p. 58.
61 For an earlier overview of the landscape, see Ebener and Betancourt (1985).
the status quo or make a more radical move remains a subject of open debate. One who argues for the latter is Hadfield (2001), who suggests that the key to achieving economic efficiency is to privatize commercial law entirely.

5.9. Fee-shifting

Under the “American rule” each party pays its own costs of litigation, whereas under the “British rule” the loser pays all. The theoretical literature on fee shifting leaves an indeterminate prediction as to the effect of a change from the American to the British rule on the probability of settlement.\(^{62}\) Cooter and Rubinfeld (1989) review that literature, suggesting that in cases on the margin between trial and settlement where plaintiffs are likely to be more optimistic about expected trial outcomes, a switch to the British rule will reduce the bargaining surplus to be gained by settling, which will increase the probability of trial. However, two forces go in the opposite direction. First, under the British rule the stakes of the case are effectively increased.\(^{63}\) To the extent that this encourages the parties to expend more effort at trial, the former effect will be muted or eliminated. Second, under the British rule the parties face greater risk, which would be expected to increase settlements.

Because the direction of the overall effect on the frequency of trials cannot be determined from theory alone, the empirical evidence will be decisive. The view among many lawyers that fewer suits will occur when the loser pays more is supported in part by evidence from experimental economics. In one study, Coursey and Stanley (1988) found a higher settlement rate when their experimental subjects decided disputes under the British rule as opposed to the American rule. However, in a more recent study, Main and Park (2000) found a different result—the British and American rules produced no difference in the frequency of settlements overall. As theory would predict, they did find that the British rule produces higher settlements in cases in which the plaintiff’s probability of success is the highest.

There is very little actual empirical and econometric evidence concerning this topic. One interesting exception is Hughes and Snyder (1995).\(^{64}\) The authors use data from a natural experiment that arose when the state of Florida switched from the American rule to the British rule for medical malpractice litigation between July 1980 and September 1985. They draw the conclusion that the British rule does tend to increase the fraction of cases settled (and the fraction that are dropped). Cases that do go to trial tend to generate more pro-plaintiff outcomes.

---

\(^{62}\) A switch to the British system is typically thought to reduce the incentive to bring a suit; see, for example, Bebchuk and Chang (1996), Polinsky and Rubinfeld (1996).

\(^{63}\) Katz (1988) simulates a 125% increase in costs.

\(^{64}\) See also Snyder and Hughes (1990).
5.10. Class actions

Class actions allow an attorney or attorneys to bring a civil lawsuit on behalf of a large number of plaintiffs. Seen from the plaintiffs’ perspective, class actions allow plaintiffs to recover for harms suffered when it would be uneconomic to bring individual lawsuits. From the defense perspective, class actions are seen as overdeterrent to the benefit of class attorneys more than the members of the class. Broader policy issues surround the overall costs of class action litigation, but also the nature of settlements (e.g., should coupon settlements be allowed?) and the extent of legal fees (how actively should courts monitor fees?).

Most empirical studies of class actions have been descriptive in nature, focusing primarily on mass tort cases. Hensler (2001) offers a recent, useful review of the literature, highlighting the results of a study of class actions conducted by the RAND Corporation’s Institute for Civil Justice. She points to the fact that tort cases actually represent somewhere between 9% and 18% of all class actions active in the 1995–96 period. Based on interviews with class action attorneys and documentation of active cases she offers a nuanced view that supports neither of the two extreme positions described previously. While rich in detail, the study leaves much open to future empirical work, since it does not attempt to make an overall policy assessment of the benefits and costs of the class action approach to litigation. Thus, the study found that average compensation collected or projected to be collected ranged from $270,000 to $840 million, with average payments per class member ranging from $6 to $1500 in consumer suits and $6400 to $100,000 in mass tort suits. Whether those “benefits” are worth the costs of litigation remains an unanswered question.

With respect to settlement behavior, it is believed agency issues make it likely that some cases will settle when it is in the interests of class members to go to trial (because the attorney can obtain a relatively high fee by settling and avoid the costs of trial). One early study, Rosenfield (1976), provides some supporting evidence based on a sample of over 100 class actions. More recent studies have not contradicted this view, but settlement has not typically been a focal point of the work. Thus, Willging’s 1996 study for the Federal Judicial Center found that a majority of certified class actions were settled, and this rate was 2 to 5 times as high as cases that contained class actions that were not settled. However, Willging did not attempt to control for case characteristics that might distinguish the value of the underlying cases.

Acknowledgements

The authors wish to thank Ashley Ensign for her helpful research assistance, and would also like to thank Eric Rasmussen and the editors for their thoughtful comments.

65 See Hensler et al. (1999), and Willging, Hooper, and Niemi (1996).
References

Contracts


of State Supreme Courts’ Bad Faith, Breach-of-Contract, Breach-of-Covenant-of-Good-Faith and Excess-
NCJ-156664.
U.S. Department of Justice, Bureau of Justice Statistics (2000a). “Contract Trials and Verdicts in Large Coun-
ties, 1996”, NCJ-179451.

**Torts: Introduction and medical malpractice**

Organization 2, 279–303.
porary Problems 5, 57–84.
Journal of Health Economics 18, 491–522.
Tort Litigation in the 1990s”, Georgia Law Review 34, 1056–1154.
641–695.
Harvard Medical Practice Study (1990). “Patients, Doctors, and Lawyers: Medical Injury, Malpractice Lit-
gation, and Patient Compensation in New York”, a report by the Harvard Medical Practice Study to the
Press, Washington, DC.


Torts: Automobile bodily injury


Torts: Products Liability


Torts: Mass torts


Torts: Punitive damages


Property


Ch. 5: Empirical Study of the Civil Justice System


Legal process


