1. INTRODUCTION

This chapter on the regulation of medical professions discusses only two of the major areas where researchers have investigated, theoretically and empirically, the impact of regulation of the medical professions, the licensing of medical professions and medical malpractice law. Although not addressed in this chapter, a wide and diverse literature on other aspects of the regulation of medical professions certainly exists (e.g., the impact of capital constraint, or Certificate of Need, regulation, see Joskow, 1981; for a general overview of the wider literature see Feldstein, 1988 and Phelps, 1992). Although the law and economics literature has made major contributions to both of these two topics, there exists very little overlap between the two areas in either theoretical or empirical studies.

The literature on professional licensing relies upon a wider literature, explaining the existence of regulation in general, for its theoretical base. From this literature, three explanations for professional licensure emerge: (1) the capture theory, which assumes that professionals “capture” regulation and use it to deter entry and increase their incomes, (2) the public interest theory, which assumes that professional licensure is used in the public interest in order to insure the quality of professional services and (3) the political economy theory, which assumes that both professional interests and public interests may simultaneously have an impact upon the existence and nature of licensing for health care professionals.

The remainder of the professional licensing literature is focused on attempts to empirically assess the validity of these three theories. Although most empirical researchers conclude that the data support the capture theory, a careful reading of their findings actually reveals that such support is inconsistent at best. Thus, although medical licensure is sometimes found to deter entry and enhance the income of medical professionals, at other times it does not have such an impact. The inconsistent support for the capture theory, combined with a few studies which directly test both the public interest theory and the political economy theory, lend support to the political economy theory. Very little additional research in this area has been published in the past decade since the publication of the first edition of the Encyclopedia of Law and Economics (Olsen, 2000).
Medical malpractice is the second, and arguably more successful, method of regulating medical professions. The theoretical literature on medical malpractice is also based upon a wider literature dealing with tort law generally rather than medical malpractice specifically. This literature focuses on the impact that malpractice liability has upon the efforts made, or precautions taken, by the medical professionals in order to avoid medical accidents which may result from their treatment of patients. As with tort law in general, the major aim of medical malpractice is, theoretically, to deter medical professionals from shirking such precautions. In the section on medical malpractice, this general theoretical literature is discussed in more detail, as are the extensions of the literature for medical malpractice, specifically.

Although the medical malpractice literature has made some original contributions to the theoretical literature, the main focus of the malpractice literature, as with the professional licensing literature, has been on empirical studies. One of the main areas where the field of law and economics has contributed has been in studying, empirically, the periodic medical malpractice “crisis” which has arisen in the past 30 to 40 years. As detailed below, during this time period the liability of medical professionals has increased dramatically and this has led to the common conclusion that the malpractice system itself is inefficient and in need of reform. The review below critically assesses the validity of this common conclusion and advises caution, given the nature of the data to be found in the literature and, especially, the common lack of a historical perspective concerning malpractice liability. In fact, one of the major themes in the literature in the past decade is to follow up on this weakness in the literature by showing the existence of a number of myths about medical malpractice generally and the periodic malpractice “crises” specifically.

The final area of concern in the literature, also addressed below, is medical malpractice reform. Naturally, given the common assumption that a crisis exists in medical malpractice, much of the literature focuses on the need for reform of the malpractice system. Some of this literature is theoretical in nature, discussing the types of reforms that would be most likely to move the system away from the crisis and towards a more efficient system. Some, perhaps a majority, focus on detailing the actual reforms that have been enacted, mostly by the US and entirely at the state level, and estimating their effectiveness in reducing the malpractice liability of medical professionals.

2. THE LICENSING OF MEDICAL PROFESSIONS

The literature on the licensing of medical professions fits into a larger literature on regulation in general and the licensing of all professionals in particular.
These literatures have a common theoretical background which explains the existence, and persistence, of regulation. A brief summary of this theoretical literature is included below in order to provide a framework for the subsequent discussion of the empirical literature on licensing medical professions. Thus, the major contribution of the literature on medical licensure has not been theoretical but empirical. The extensive literature which investigates the impact of licensure upon markets for medical services is synthesized below. Finally, conclusions about findings gleaned from the empirical literature and directions for future research are given in the final section.

2.1 Theoretical Background

Three general theories have been used to explain the existence of regulations limiting professional licensure of the medical professions: the capture theory, the public interest theory, and a more general, political economy theory. Although these theories are well known in the economics literature, a brief overview of each is included both for background and to clarify the testable hypotheses flowing from each theory.

In its simplest form, the capture theory is a straightforward application of self-interest. Medical professionals “capture” the regulations governing licensure and structure them to limit the supply of medical professionals and thereby increase their incomes (Coase, 1974; Friedman, 1962; Moore, 1961; Posner, 1974; and Stigler, 1971). As a result, the capture theory predicts that licensure should decrease the supply of medical professionals, increase the prices charged by medical professionals, and increase existing medical professionals’ incomes.

Of course, regulatory capture may be attempted by coalitions of professionals within a given profession to the detriment of both the public and non-coalition professionals. For example, regulations may be structured to benefit physicians in certain specialties. In these cases, incomes of only coalition professionals would be expected to rise.

In contrast, the public interest theory suggests that professional licensure occurs due to some market “failure” and that its intent is to increase societal welfare. Unlike the capture theory, the public interest theory does not yield clear-cut, tractable hypotheses that can be readily tested. For example, predictions from the public interest theory vary depending upon the type of market failure. Thus, the first step is to accurately identify the relevant market failure being addressed. Researchers have commonly presumed that consumers of professional services, due to the complex nature of the service and uncertainty about the efficacy of competent service, lack complete information about the quality of such services (e.g., Arrow, 1963). Hence, the public interest theory asserts that professional licensing corrects this market deficiency by ensuring that medical professionals are of a sufficiently high standard and quality.
However, even if such informational asymmetries exist for professional services, the public interest theory still does not yield clean, testable hypotheses. For example, Leland (1979) shows that the impact of quality standards (i.e., professional licensing) is dependent upon the cost function for quality. If higher quality applicants can meet the quality standard at lower cost, quality will rise, as will market price and professional income. However, if higher quality applicants have a higher cost of meeting the standards, quality, prices, and income will all fall. Thus, to be able to predict the impact of professional licensure one must also know its impact on quality. Unfortunately, data on quality are notoriously difficult to obtain. As a result, most researchers interpret the lack of anti-competitive effects as support for the public interest theory (e.g., Holen, 1965; Maurizi, 1974; Shepard, 1978; and Noether, 1986). To be clear, anti-competitive effects are those which are predicted by capture theory – a decreased supply of professionals, higher prices, and higher incomes for professionals.

In contrast to both capture and public interest theories, the political economy theory entertains the possibility that both the public and medical professionals have an impact on the existence and the form of medical professional licensure regulations (Bentley, 1967; Becker, 1983; Peltzman, 1976, and Stigler, 1971). Thus, the political economy theory is basically a theory of checks and balances. Since all have an impact on the regulatory outcome, medical professionals cannot monolithically shape regulations merely to suit their self-interest; their ability to do so is tempered by the competing self-interest of the “public.” In fact, Becker (1983) demonstrates that regulators will have a predisposition to enact regulations that are in the public’s (society’s) interest. However, this predisposition is tempered by other factors such as the dispersion or concentration of the affected parties. Thus, the political economy theory predicts that medical professionals will sometimes, but not always, be able to use licensure to limit supply and increase prices and incomes.

2.2 Empirical Studies

Before summarizing the empirical literature, one must briefly discuss the nature of medical licensure regulation and how licensure is measured in the empirical studies. There are several common requirements before a medical professional can receive a license to practice. For example, it is common to require that professionals obtain a given education at an approved educational facility. Likewise, many medical professions require candidates for licensure to pass either state or national licensure exams. As can be seen below in the discussion of the empirical literature, the difficulty of the licensure exam as measured by its pass rate (or failure rate) in a given jurisdiction is commonly
used to evaluate the impact of the exam requirement on market conditions. Educational and exam requirements are the most common licensure requirements. Some jurisdictions, however, waive educational and exam requirements for certain professionals, most commonly those who have obtained a license to practice in another jurisdiction. Some empirical studies also estimate the impact that such rules, known as reciprocity, have on market conditions. Licensure requirements, of course, vary by jurisdiction and by profession.

Though beyond the scope of this chapter, a full examination of the history of licensing for medical professionals can provide insight into the issues of primary concern in the literature discussed in this chapter. For example, licensing for health care professionals, as with many other professions, has most often either been proposed by the regulated professionals themselves or the licensing has had the support of the regulated professionals. In fact, physician licensure only became truly effective and widely applied jurisdictionally with the formation from the late 1800s of effective national and state medical associations who could push for licensure (Shryock, 1967; Starr, 1982; Frech, 1996, Law and Kim, 2005). Such an impetus for licensing, though most often combined with regulated professionals publicly arguing that licensing served the public interest by protecting the public against low quality professionals, lends implicit support to the capture theory discussed above.

For a summary of the history of licensure in the nursing profession, see White (1980, 1983, 1987). The empirical literature also discusses current licensing regulations and the history of licensing regulations for other medical professions such as dentists (Shepard, 1978; Boulier, 1980), physicians (Kessel, 1958; Shryock, 1967; Leffler, 1978; Rizzo and Zeckhauser, 1990; Frech, 1974; Frech, 1996), optometrists (Benham and Benham, 1975) and physical therapists (Sass and Nichols, 1996). Other studies discuss licensing in general or for a number of different professions (Moore, 1961; Holen, 1965; Starr, 1982; Pashigian, 1979; Frech, 1996).

A careful review of the literature yields no consistent picture of the impact that medical licensure has on income, prices, supply, or the quality of medical professionals. In support of this conclusion, Garoupa’s (2004) study compares European to US licensing regulations and finds that its impact varies greatly between the different countries considered. Earlier US studies in this area focused on estimating the return to medical education under the assumption that higher than average returns to medical education, as opposed to other types of education, would be an indication of monopoly rents obtained through the capture of licensure. Friedman and Kuznets (1945), for example, compared the incomes of physicians to dentists, and found that physicians had average incomes 32.5 percent greater than dentists. According to their calculations, only approximately half of this increased income could be a result of extra training requirements for physicians. As a result, they concluded that barriers
to entry must explain the remainder of the differential returns. More recently, Law and Kim (2005) have confirmed Friedman and Kuznets’ conclusions using a historical data set from the 1880s to the 1930s. However, they found no statistically significant impact of licensing on physician gross income or net income. Early studies by Sloan (1970) and Fein and Weber (1971) also found higher than average returns to medical education and reached similar conclusions.

However, Lindsay (1971, 1973) demonstrated that these empirical studies overestimated the returns to medical education. Lindsay noted that medical professionals were characterized not only by higher than average income but also by higher than average weekly hours worked. For example, during the period of the Fein and Weber (1971) study, 1966, physicians worked an average of 62 hours per week. As a result, estimates of the returns to medical education must be reduced to offset the disutility of working longer hours. Lindsay (1973) corrected for the hours worked bias for the time periods covered by all three of the earlier studies and found, as would be expected, much smaller differential returns to medical education. In some cases, he found negative differential returns, dependent upon the discount rate chosen, to medical education. Hence, Lindsay’s results seriously question the conclusions of earlier studies that entry barriers provided by, among other factors, medical licensure had led to higher than average returns.

Leffler (1978) also noted that the differential returns to medical education were overestimated not only because of an hours worked bias but also because of the increased expected mortality of physicians, progressive income taxation, and “differential probabilities of conscription taxation” (Leffler, 1978, p. 167). After adjusting for these factors, Leffler found much lower differential returns to medical education. Leffler did find positive differential returns, again dependent upon the discount rate, in some years, particularly the late 1960s and early 1970s. However, he attributed these returns to transitory demand shocks rather than entry barriers.

Other studies have estimated hours adjusted rates of return to medical education with more recent data (Burstein and Cromwell, 1985; Marder and Willke, 1991) and have found positive differential rates of return. For example, Burstein and Cromwell (1985) found internal rates of return, hours adjusted, equaling about 12 percent for physicians and dentists from 1970 to 1980. Likewise, Marder and Willke (1991) estimated internal rates of return, hours adjusted, by physician specialty in 1987 and found that such returns varied from a low of 1.5 percent for Pediatrics to a high of 58 percent for Pathology (where hours adjustment increased the rate of return) and 3.7 percent for General Practice. Thus, even though rates of return to medical education are not as high as estimated originally, substantial returns do exist, especially for some medical specialties.
In addition to studies which focused on estimating the differential return to medical education, other studies have estimated the direct impact that licensure has had on the medical professional’s income. For example, early studies on professional licensure by Holen (1965), Benham et al. (1968) and Maurizi (1974) suggested that licensure (as measured by licensure exam difficulty) was used to increase the incomes of professionals. The support for this conclusion was, however, mixed in these studies. For example, Benham et al. found licensure to have a positive and significant impact on dentists’ income. Licensure, however, was found to have a negative, and sometimes significant, impact on physicians’ incomes. In addition, the studies used small samples of older aggregate data, rather than recent individual data. Further, neither Holen nor Maurizi attempted to control for the possibility that other demand or supply variables had a significant impact on professional income.

Considerable overlap exists in many studies estimating the impact of professional licensure. Both Holen and Maurizi, for example, tested the impact of licensure regulations on multiple professions, including both medical and non-medical professions. Maurizi tested the impact of regulations on 18 different professions, including eight medical professions (chiropractors, dentists, registered nurses, optometrists, osteopaths, pharmacists, and physicians).

Additional studies of the impact of medical licensure on professional income also yielded mixed results. For example, Shepard (1978) found that lack of reciprocity between jurisdictions significantly limited the number of dentists and, as a result, increased average income for dentists. Shepard, however, used state level aggregate data for a single year (1970), which limits the validity of the results. White (1980) found that mandatory licensing for nurses had a small, but positive, impact on nurses’ incomes when first introduced in the 1950s. However, in later years, after licensing had had more time to affect incomes, White found that mandatory licensing had no significant impact on nurses’ incomes. Pashigian (1980) found that licensing in a number of occupations, including physicians, dentists, optometrists, pharmacists, nurses and other allied health occupations, restricted the mobility of licensed professionals, all else equal, but did not significantly raise incomes (for similar work, see Pashigian, 1979).

Noether (1986) found that entry restrictions on medical school graduates from non-US medical schools increased average physician income during the early 1960s. However, Noether’s data indicate that entry restrictions have not been effective since 1965 in increasing physicians’ incomes. She attributes this result to the increased ability of foreign medical school graduates to enter the US market.

In a study of physical therapy licensure, Sass and Nichols (1996) found that licensure actually caused a decrease in professional wages and, presumably, in...
income as well. Licensure in the Sass and Nichols study is somewhat different from that measured in most other empirical studies of medical licensure. Rather than focusing on licensure conditions which limit the entry of professionals, Sass and Nichols focus on states with direct access laws (i.e., laws which allow physical therapists to treat patients without referrals from physicians). Although physical therapists have lobbied in favor of direct access laws, Sass and Nichols found that direct access has the impact of lowering therapist wages and, presumably, income as well.

Leffler (1978) found that physicians in states which rejected national physician licensure exams in favor of state licensure exams had higher average incomes. Leffler’s state level data comes from the 1960s and early 1970s. At that time, 41 states allowed physician candidates for licensure to take either a standardized national exam or an exam specific to the given state. Nine states, however, did not allow candidates the option of taking the national exam but required candidates to pass the state exam for licensure. Requiring the state exam for licensure reduced the mobility of professionals from other states and, according to capture theory, would be expected to increase physician incomes. As a result, Leffler concluded that some evidence existed that local state coalitions of physicians used licensure in order to increase incomes. The fact that only nine of the 50 states required candidates to take the state exam indicates, however, that physician capture of licensure occurred in only a few states. Further, Leffler found direct evidence that states that accepted the national exams did so as a result of consumer, rather than physician, demand for licensure.

Some evidence does exist indicating that medical licensure has a positive impact on prices charged by medical professionals. Kessel (1958) examined pricing of physician services and found evidence that suggested that physicians practiced price discrimination. More importantly, Kessel provided evidence that local medical societies, who controlled licensure, used licensure to force local physicians to keep prices high and comply with price discrimination practices.

The major problem with Kessel’s study is his reliance upon ad hoc data. As a result, Kessel could only detail efforts by some local medical societies to force physicians to comply with local pricing rules. In fact, Kessel’s study focused on opposition by local medical societies to pre-paid health plans and the physicians who contracted with such plans. He argued that opposition to pre-paid plans resulted because they did not generally allow price discrimination by patient income. Kessel’s ad hoc empirical evidence, then, focused on demonstrating that local medical societies often denied licensure, or attempted to deny licensure, to physicians affiliated with such plans. Given the nature of Kessel’s ad hoc data, though, he could not estimate whether or not anti-competitive efforts by local medical societies were actually successful in raising prices.
or promoting price discrimination. In fact, the ad hoc data which Kessel used demonstrated that, in some instances, efforts to deny licenses to non-complying physicians failed.

Studies by Benham and Benham (1975) and Haas-Wilson (1986) on eyeglasses and Shepard (1978) and Boulier (1980) on dental services provided some evidence that professional licensure increases prices. These studies do have some weaknesses. Only Haas-Wilson and Boulier used individual data. Haas-Wilson’s results were strong only when she considered the cumulative effects of several regulations. Boulier did not measure licensing restrictions directly, but, rather, found state variations in individual fees which could not be explained by other variables included as controls in the regressions. The possibility of omitted variables limits the validity of his study. Shepard’s study is the most convincing but suffers because he looked at the impact that licensing has on average prices rather than individual dentist’s prices and because he uses only state aggregate, rather than individual, data.

Next, consider the impact that medical licensure has on the quantity of professionals in a given market. As noted above, both Shepard (1978) and Noether (1986) found that licensure reduced the number of professionals in a market. Shepard with dentists and Noether with physicians. Noether found that entry restrictions significantly limited the ability of foreign medical school graduates to enter the US market, which, in turn, positively affected physician incomes in the US. However, she also found that the effectiveness of such entry restrictions has deteriorated over time to the point where their impact was largely non-existent by the early 1980s. White (1980), in his study of licensing for nurses, found that licensure either had a positive impact on employment of nurses, as it did in the early years of licensure, or else had no significant impact on employment, as it did in later years. Svorny (1987), using aggregated state level data from 1965, found that licensure reduced the number of physicians per capita. Law and Kim (2005) found that physician licensing reduced the entry of physicians from 1880 to 1930. However, the only licensing requirements found to have a statistically significant negative impact on entry were those related to education.

Finally, consider the impact that professional licensure has on the quality of professional services. Studies of quality almost universally suffer from one overwhelming weakness: quality is difficult to measure. Regardless of this practical difficulty, several studies have attempted to estimate the impact that professional licensure has on the quality of professional services. Haas-Wilson (1986) generally found that regulations in optometry had no significant impact on quality as measured by the thoroughness of the eye exam.

Carroll and Gaston (1981) found scattered support for their conclusion that licensure resulted in decreased quality for several professions. Their results were strongest for dentistry, although the variable they used to measure the
quality of a dentist’s services – how long a patient must wait to receive an appointment rather than by a direct measure of the effectiveness of the dentist’s services – is questionable at best. In contrast, Feldman and Begun (1985) considered the impact of a number of different licensing regulations for optometry and found that regulation increased the quality of services as measured by the length of the eye exam and its thoroughness. Finally, Law and Kim (2005) found decreased mortality rates for some, but not all, procedures, as states first adopted physician licensing in the 1880s to the 1930s. These results provide some evidence that initial licensing regulations had some positive impact on quality.

2.3 Conclusions

The review of the empirical literature reveals that medical licensure does, for certain medical professions, in some mostly US jurisdictions (primarily US states), or during some periods of time, have the anti-competitive impacts predicted by the capture theory. At times and for certain professions, medical licensure limits the number of professionals in a market, with resultant upward pressure upon prices of professional services and professional incomes. It is not surprising, then, that many of the empirical studies on medical licensure have concluded that their empirical results tend to support the capture theory.

However, taking a broader overview of the entire empirical literature, one finds that at other times or for other medical professions, anti-competitive effects are not present. The same inconsistent picture holds when examining the impact that medical licensure has upon the quality of medical services. Sometimes, licensure increases quality, as predicted by the public interest theory, but at other times or for other medical professions it does not.

The fact that no consistent picture of the impact that professional licensure has on supply, prices, income or quality of medical professionals emerges when reviewing the empirical literature seriously weakens support for both the capture theory and the public interest theory. The inconsistent impact of medical licensure actually lends support to the political economy theory of licensure where both medical professionals and the public are expected to have an impact on licensure. The political economy theory predicts, as noted above, that professionals will sometimes, but not always, have the ability to capture licensure to serve their own best interests. At other times, or for other professions, licensure will be used to serve the best interests of the public.

Although most empirical studies of medical licensure conclude that their results support the capture theory, a few studies have taken an explicit look at the political economy theory. For example Leffler (1978), in his study of physician licensure, sets up a model to explicitly test the political economy theory. Leffler found evidence supporting consumer demand for physician
licensure as well as significant pressure by physicians using licensure in an attempt to raise their incomes. Again, that neither physicians nor consumers are always successful in their attempts to use licensure gives support to the political economy theory of licensure. Likewise, Graddy (1991) found support for public interest and legislative forces, as well as interest group pressure, in her study of state occupational regulations.

Thus, the most serious failing of the medical licensure literature is the axiomatic approach of many researchers – assuming that the capture theory is correct and, not surprisingly, concluding that their own evidence, even though it is often only weakly supportive, supports the simple capture theory. As can be seen, a wealth of empirical studies have been done which look at the impact of licensure for a number of different medical professions. Only a few studies, though, have been done which seriously consider, or test for, the possibility that licensure has a more complex explanation. The major area where future research, mostly empirical, would be most profitable lies in testing these more complex explanations for medical licensure.

3. MEDICAL MALPRACTICE

The previous section of this chapter illustrates that the direct regulation of medical professions via licensing requirements has only inconsistently improved quality of care. At times, it does appear that professional licensing in medicine is in the public interest, but at other times evidence exists that such regulation actually is beneficial to the industry being regulated rather than the public. However, professional licensing is not the only method used to regulate the quality of care in medicine. Medical malpractice law is another, and arguably a substantively more effective, method of regulating the quality of care in medical professions.

In the past 40 years, medical malpractice has consistently been of interest to a wide variety of groups, including but not limited to the law and economics literature. The intense interest in medical malpractice in the past three decades has been a result of the widespread perception that the malpractice system, previously working smoothly, experienced periodic crises beginning in the 1960s and extending into each decade since (e.g., Danzon, 1985b and 1988; Sloan et al. 1991; Weiler, 1991; Baker, 2005). The widespread perception of a malpractice system in crisis was fueled by large increases in physician liability beginning in the 1960s and, perhaps more importantly, periodic disruptions in the malpractice insurance market beginning at the same time. Besides large increases in malpractice insurance premiums, unsurprising considering the increases in physician liability, the perception of a crisis was furthered when in some US states malpractice insurance became unavailable.
at any premium. The perceived crisis in medical malpractice has generated an extensive literature in a number of diverse areas. The major areas where law and economics has contributed to the literature on medical malpractice are discussed in more detail in the following sections. For example, the law and economics literature has made a large contribution in developing the theoretical justification for the existence of medical malpractice law.

However, empirical studies have been, perhaps, the most significant contribution of law and economics to the medical malpractice literature. These studies have focused on three main areas. First, a number of studies detail the nature and extent of the “crisis” in medical malpractice liability and the malpractice insurance market. Second, given the common conclusion that medical malpractice is in crisis, a number of studies focus on medical malpractice reform. Some of these studies are theoretical in nature, arguing for a particular type of reform as that best able to correct perceived deficiencies in the medical malpractice system. Others attempt to measure the impact that US legislative reforms passed at the state level have had upon the medical malpractice system. Most, although not all, of the focus in the literature has been on US medical malpractice law, which leaves other malpractice systems as fruitful areas for future research. Third, more recently, a new literature has arisen that empirically tests a number of the key assumptions regarding the assumed underlying causes of the periodic medical malpractice crises. The literature has tested many of the assumptions underlying the assumption that the malpractice system was in crisis and often found such assumptions to be more mythical than factual.

3.1 The Theory of Medical Malpractice

The theoretical justification for medical malpractice law fits into a larger law and economics literature discussing the existence and economic efficiency of tort law. In addition to this more general theoretical framework, the literature has also developed some theory specific to medical malpractice.

Medical malpractice law, and tort law generally, hold health care providers to a court-determined legal standard of care which, if violated, establishes the negligence of the provider. Negligence will, in turn, establish the liability of the physician, or other health care provider, for any resultant injuries suffered by the patient (see Keeton et al., 1989, for a complete discussion of the law of torts).

Tort law, generally, and medical malpractice law, specifically, serve two primary purposes. First, they compensate injured parties for losses caused by negligent accidents. Second, the threat of liability for negligence induces potential injurers to take precautions against potential harm (Brown, 1973; Landes and Posner, 1980 and 1984; Cooter and Ulen, 1986; and Posner,
The medical malpractice literature tends to focus on the incentives created by medical malpractice law, known as the deterrence effect, rather than upon the compensation of negligent accidents, although, certainly, exceptions exist (e.g., Shavell, 1978). The larger law and economics literature does address the issue of compensation as it is related to involuntary insurance against adverse outcomes. That is, higher average compensation levels will, all else equal, induce potential injurers such as physicians to increase the price of their services to all patients. The increased compensation is, in effect, an implicit insurance contract against the risk of potential malpractice injuries. This larger literature generally argues that from the point of view of optimal insurance, which is defined to be that level of insurance which a fully informed consumer would choose to buy himself, full compensation levels lead to higher levels of implicit insurance than is optimal (see Epstein, 1980; Priest, 1985, 1987; Schwartz, 1988; and Viscusi, 1991).

With respect to deterrence, the theoretical literature attempts to determine whether or not negligence, or other alternative liability rules, will lead to optimal or efficient deterrence. Optimal deterrence occurs when the liability rule gives each party (i.e., both health care provider and patient) an incentive to take actions, or precaution, which minimize the total costs of the medical injuries, including the expected accident costs of the injuries themselves, the costs of preventing such injuries, and the administration costs associated with the malpractice system (Calabresi, 1970).

Brown (1973) and others (e.g., Landes and Posner, 1980 and 1984; Cooter and Ulen, 1986; and Posner, 1986) have shown that tort liability rules can be used to force both parties (i.e., in medical malpractice, both provider and patient can take actions to avoid harm from treatment) to consider the full potential loss and, hence, to take the optimal level of precaution. Unfortunately, this result is only valid under very limiting conditions; the courts and both parties must have full and costless information. When information is costly to obtain and, hence, incomplete, tort law does not necessarily result in optimal deterrence (Cooter and Ulen, 1986; Haddock and Curran, 1985; Kolstad, Ulen, and Johnson, 1990; Shavell, 1987; and White 1989).

Even though the theoretical literature on medical malpractice draws upon the general theoretical background of tort law, described above, the medical malpractice literature goes beyond this general theoretical framework, although not always by using formal, mathematical models of deterrence. For example, a number of researchers have addressed the issue of which liability rule, out of all possible liability rules, results in optimal deterrence for cases of medical malpractice (e.g., Shavell, 1978; Danzon, 1985b). Currently, as noted above, medical malpractice uses some form of the negligence liability rule, where the courts establish a legal standard of precaution against which a provider’s actual precaution is measured in order to determine negligence and
liability. Alternatives to a negligence liability rule, such as a strict liability rule or a no-fault rule, could be used to determine liability (see Shavell, 1978; Danzon, 1985b; Epstein, 1978, 1988; Weiler, 1991).

In contrast to a negligence rule, a strict liability rule does not compare a defendant’s level of precaution to a legal standard. Rather, as long as the accident or injury was caused by the medical treatment, the defendant is found liable for the costs of such accidents. Theoretically, the major advantage of strict liability is that it reduces the costs of the legal system because only causation must be demonstrated rather than, as in the case of negligence rules, both causation and negligence. However, determining causation for medical injuries is problematic as adverse medical outcomes routinely happen because of prior medical conditions. In medical malpractice, unlike other torts, causation is often determined to exist when an adverse outcome occurs and the defendant is found negligent. Hence, a finding of negligence is often required to prove causation which obviates the cost advantages of strict liability.

Similar problems exist relative to no-fault proposals for medical malpractice because no-fault rules generally require proof of causation as well (Shavell, 1978; Epstein, 1978, 1988; Danzon, 1985a).

Thus, medical malpractice gives health care providers an incentive to avoid negligent injuries caused by medical treatment. Under ideal conditions, medical malpractice results in optimal deterrence. However, since conditions are not, realistically, ideal, some thought, as noted above, has been given to alternatives to the negligence standard set up by medical malpractice, such as strict liability and no-fault rules. However, there is another, useful, way to look at medical malpractice. For example, Posner (1986, 1981) has proposed that the law, especially common law, can be best understood as a method of achieving societal efficiency (i.e., maximizing the wealth of society). Under Posner’s efficiency theory of the law, legal rules do not necessarily result in perfectly optimal, or first-best, outcomes. Rather, the law is seen as a system for achieving the best possible, or second-best, outcomes given the constraints faced in the real world (e.g., costly and, hence, incomplete information).

Given the efficiency theory of the law, it is useful to look at medical malpractice as an efficient, albeit imperfect, system. Somewhat surprisingly, few researchers have attempted to examine medical malpractice from this viewpoint, especially in any formal sense. A number of researchers have, though, loosely noted that medical malpractice may either be efficient or not as inefficient as its harshest critics have claimed (e.g., Frech, 1978). Even so, a number of these researchers continue by proposing reforms that, they suggest, will improve the malpractice system, implying, at the very least, that malpractice has not yet reached its second-best optimum (e.g., Danzon, 1985b). Thus, one of the flaws in the theoretical literature on medical malpractice remains an insistence that medical malpractice law is inefficient. As
shown in the next section, the persistence of the view that malpractice is inefficient can be, for the most part, traced to the widespread perception in the literature that the medical malpractice system is in "crisis."

3.2 The Medical Malpractice Crisis

Perhaps more than any other event, significant changes in the medical malpractice insurance market during the past few decades have prompted the view of a malpractice system in crisis. For example, available evidence indicates that during most of the 1950s the market for medical malpractice insurance was quite stable, with malpractice insurance being so unimportant that insurance agents typically sold medical malpractice policies to physicians along with their home and auto insurance (Sloan et al., 1991).

Beginning in the 1960s, however, malpractice insurance premiums rose dramatically, over 500 percent in the 1960s (Sloan et al., 1991; Posner 1986) and by the mid 1970s a full blown “crisis” was thought to have developed (Danzon, 1985b; Sloan et al., 1991; Robinson, 1986b). Not only did insurance premiums rise dramatically, but in some states, most notably New York and California, malpractice insurance was not available at any price as insurance carriers left the states. Increases in malpractice insurance premiums continued into the 1980s. By some estimates, premiums increased by as much as 45 percent from 1982 to 1984 (US GAO, 1986). Of more importance, medical malpractice premiums, which in the 1950s represented less than 1 percent of a physician’s total costs (Sloan et al., 1991), had risen to 9 percent of their total costs by 1984 (US GAO, 1986).

Both malpractice liability and premiums were more stable during the 1990s. However, more recently, medical malpractice premiums have increased at substantial rates again (Thorpe, 2004; Baker, 2005). This more recent “crisis” has also been typified by a reduction in the number of insurance firms that offer medical malpractice coverage (Thorpe, 2004 and American Medical Association (AMA), 2002). Thus, the data seem to suggest three separate medical malpractice “crises”, the crisis occurring in the 1960s and into the 1970s, a second crisis during the 1980s, and a third crisis beginning in 2000 or 2001 (see, e.g., Thorpe, 2004; AMA, 2002 and Kilgore et al., 2006).

The empirical medical malpractice literature gives insight into why malpractice premiums rose at such dramatic rates during the first two of these time periods (i.e., the 1970s and 1980s). Essentially, as with all insurance premiums, the root cause of increased premiums can be traced to increased average benefits paid by insurers. Three variables affect the average benefits paid by medical malpractice insurers. The frequency of malpractice claims measures how often, usually on a per capita or per physician basis, malpractice claims are being filed. The probability of malpractice claims measures the
likelihood, most often measured by the ratio of paid claims to total claims for a period, that the physician will lose a given malpractice case. Malpractice severity simply measures the average size of the awards paid by defendants and their insurance companies for cases they lose. As malpractice frequency, probability, or severity rise, the cost of providing insurance coverage increases and companies react by increasing malpractice insurance premiums (Olsen, 1996).

A good part of the empirical malpractice literature has focused on estimating these three variables. The literature demonstrates that, not surprisingly, during the same decades that malpractice insurance premiums rose dramatically, so did malpractice frequency, probability and severity. For example, although claim frequency remained stable from the 1950s through the early 1960s (AMA, 1963), by 1968 claim frequency had risen 76 percent from its earlier level (Danzon, 1985b). From the mid 1960s to the mid to late 1980s, claim frequency rose on average 7 to 10 percent annually. In general, however, claim frequency showed modest increases in the 1960s, rapid increases in the early 1970s, a leveling off in the mid 1970s, and more modest increases in the late 1970s and early 1980s (Danzon, 1985b; Olsen, 1996; US Department of Health and Human Services, 1987).

The available evidence, although somewhat sparse, indicates that malpractice probability also increased during this same time period. For example, some researchers estimate that plaintiffs won only 24 percent of malpractice cases in the mid 1960s, which increased to as much as 53 percent by the mid 1980s (Peterson, 1987). Although Peterson’s estimates of malpractice probability only used data gathered from a few metropolitan areas, other estimates agree that by the mid 1980s malpractice probability had risen to approximately 50 percent (US GAO, 1987; Weiler, 1991). However, more recent estimates suggest that malpractice probability remains under 50 percent, perhaps as low as 30 percent on average (Daniels and Martin, 1995).

One of the major areas of concern in the literature has been the large increases in the average size of paid malpractice claims (malpractice severity) over the past four decades. From 1960 to 1984, for example, average malpractice awards in some areas rose from 10 to 14 percent annually on average (Peterson, 1987; also see Weiler, 1991 and Danzon, 1985b). Although average awards grew at rates exceeding 10 percent annually, median awards grew at much more modest rates, as low as 4 to 5 percent annually (Peterson 1987; Weiler, 1991; Danzon, 1985b). Lower growth rates for median, as compared to average, awards indicate that much of the increase in average awards is due to a small number of cases with extremely large awards, while most cases had much more modest increases.

The major difference between the most recent medical malpractice crisis and earlier crises is that other factors besides physician liability as measured
by malpractice frequency, probability, and severity are also important (Thorpe, 2004; Baker, 2005). For example, economic cycles in the insurance industry itself are often thought to have had a large impact on recent medical malpractice premiums (Thorpe, 2004; Baker, 2004). A combination of relatively high returns on investment and lower than anticipated claims payments during the 1990s at least partially explain stable premiums during that time period. However, more recently, the investment return for malpractice firms has decreased by 25 percent (Thorpe, 2004), which in turn led to increased premiums.

Although physician liability has increased slightly more recently, increases in physician liability have either been more modest, been explained by underlying changes in the economic costs of adverse health events, or in some states even been mostly stable. Missouri, for example, after seeing mostly stable claims rates during the early 1990s, experienced declining rates thereafter, including during the most recent premium crisis (Missouri Department of Insurance (MDI), 2003). During the same time, Missouri also saw mostly stable inflation adjusted malpractice severity, albeit with a brief spike in 2000. Nonetheless, Missouri saw a large decrease in the availability of medical malpractice insurance by 2001, with a number of insurance companies withdrawing who represented over 30 percent of the market (MDI, 2003). A large literature tends to find that nationwide increases in malpractice severity have been mostly explained by increases in economic losses (e.g., Seabury et al., 2004; Hyman and Silver, 2005; Baker, 2005).

In addition, briefly compare the changes in malpractice frequency, probability, and severity noted above with changes in other areas of personal injury litigation. For the most part, only product liability consistently demonstrated increases in frequency, probability and severity comparable to those discussed above for malpractice. For example, the number of product liability cases rose 16 percent on an average annual basis from 1975 to 1989 (Viscusi, 1991). Though Viscusi’s data is expressed as the number of claims filed, rather than as a frequency, it still gives an indication that product liability claims rose at large rates during the 1970s and 1980s. In support of this view, Peterson’s (1987) data indicates that product liability cases, as a percentage of total cases, have increased from 3 to 5 percent annually from 1960–64 to 1980–84. During the same time period, malpractice cases as a percentage of total cases rose from 2 to 5 percent. Other types of personal injury cases actually fell relative to total cases (e.g., work injury and auto injury cases). Danzon’s (1985a) data from the 1970s has comparable findings as well.

Similar to the frequency of cases, Peterson (1987) finds that product liability severity is comparable to malpractice severity, rising at fairly large rates (6 to 11 percent) annually from the mid 1960s to the mid 1980s, while average
severity for other types of personal injury cases was much more modest (e.g., annual increases of 2 to 5 percent for work injury claims; 4 to 5 percent for auto injury claims). Peterson (1987) finds similar trends for probability in these other types of injury claims.

The above empirical evidence is gathered from US data. The literature does, however, give evidence that similar trends for medical malpractice frequency, probability and severity exist in other developed countries. For example, Dewees et al. (1991), compared the US to Canada and found a number of similarities. Just as with the US, Canada experienced relatively stable malpractice frequency and severity during the 1950s. Thereafter, malpractice frequency and severity rose at much higher rates. From 1971 to 1990 average annual growth rates in Canadian frequency and severity were 6.1 percent and 9.6 percent respectively. Although Canadian growth rates are similar to US growth rates, US malpractice frequency during these years was consistently eight to ten times as large as Canadian malpractice frequency (Dewees et al., 1991). However, the average size of malpractice awards was fairly similar during this time period in the two countries, with a few years where malpractice severity was actually higher in Canada than in the US (Dewees et al., 1991).

Likewise, a similar trend was found in the United Kingdom where both severity and frequency rose by about 17 percent on average annually from the mid 1970s to the mid to late 1980s. Unfortunately, the United Kingdom data presented by Dewees et al. only allow comparison with their US and Canadian data for growth rates, but not for levels, of malpractice frequency and severity.

Danzon (1990) had similar findings in her comparison of the US to Canada, Australia, and the United Kingdom. In fact, Danzon found that growth rates in malpractice severity were substantially higher, on an average annual basis, in the UK and Canada than in the US. However, she found similar growth rates in malpractice frequency (averaging about 10 percent annually) in all three countries. However, US levels of malpractice frequency were five to six times that of Canadian malpractice frequency, even though, as with Dewees et al. (1991), levels of average severity in the two countries were quite similar. She also found that UK levels of malpractice frequency were one fifth to two thirds of the US levels, dependent upon the region of the country.

Thus, the widespread belief in the existence of periodic medical malpractice crises hinges upon two crucial empirical observations. First, it hinges upon the view that, prior to the 1960s, the medical malpractice system was working smoothly with few cases and, more importantly, no significant growth in the frequency, severity or probability of medical malpractice (e.g., Danzon, 1985b at p. 2). Second, the belief in the malpractice crisis also relies upon the well-established facts, as detailed above, that all of these measures
increased dramatically after the mid 1960s and that such growth continued in the 1970s and 1980s and, more recently, in the early 2000s.

Although the latter view is well-established empirically, the former is not. Olsen (1996) uses available historical data to demonstrate that medical malpractice cases were not rare events prior to the 1950s as has been assumed by most studies of medical malpractice. More importantly, Olsen (1996) demonstrates that growth rates in malpractice frequency and severity over most of the US history from 1820 on were comparable to more recent growth rates (see also DeVille, 1990; Sandor, 1957; Sadusk, 1956). Hence, these data suggest that if a medical malpractice crisis does exist, it has been an ongoing crisis for more than 150 years. Of course, the possibility exists that there is no malpractice crisis, but that the malpractice system is simply responding to changing variables in an efficient manner. Frech (1978) identified this as a possibility in his early commentary on medical malpractice. Few researchers, as noted in the previous section, take into account the possibility that the malpractice system is working efficiently. The failure even to consider, let alone investigate, this possibility has been the most glaring deficiency in the medical malpractice literature.

This deficiency has begun to be addressed in the literature in the last decade. For example, Olsen (1997) considers the possibility that the malpractice system is efficient, with empirical testing of that hypothesis. Similar to Grady (1988), Olsen posits that increased technology and, especially, the increased physician productivity (i.e., an increased ability of physicians to avoid medical injuries) that has historically been linked with increased technology, explains much of the historical increases of medical malpractice liability. In fact, Olsen (1997) finds that such increased physician productivity has historically been correlated with increased liability, as would be efficient according to standard tort models (according to the deterrence hypothesis discussed above, it is efficient to give more productive physicians, those who can have a larger impact on medical injuries, per unit of input, a correspondingly larger incentive to avoid shirking on the use of those inputs).

Another possibility worthy of mention, though rarely considered in the literature, which might explain the medical malpractice “crisis” is that the low levels of physician liability existing prior to the 1960s might actually have been too low, resulting from physician capture of licensure and the consequent use of licensure to deny expert witness testimony in malpractice cases. A number of researchers detail the emergence of this phenomenon (Kessel, 1958; Starr, 1982; DeVille, 1990), which made it much more difficult for plaintiffs to prove malpractice against physicians. However, beginning in the 1960s it became much easier to obtain expert witness testimony against plaintiff physicians, partially due to the erosion of the locality rule (which limited expert witnesses to those practicing in the plaintiff physician’s locale, where
local medical societies had more control over physicians) and partially due to erosions in the ability of local medical societies to discipline such “unethical” practices as the offering of expert witnesses (see Danzon, 1985b; Weiler, 1991 and Olsen, 1997).

3.3 Medical Malpractice Reform

Regardless of whether a malpractice crisis actually existed or not, the widespread perception of a crisis led to public policy pressure (which continues to the present) to reform medical malpractice law in the US (Sloan et al., 1991; US Congress – Senate 1969). Although a near consensus existed regarding the need for reform, no consensus existed on the appropriate nature of medical malpractice reform. For example, some in the law and economics literature have proposed a move toward a contractual basis for medical malpractice liability (Epstein, 1976, 1978, 1986; Robinson, 1986a; and Havighurst 1983, 1986). Others have suggested that medical malpractice should adopt a no-fault liability system similar to worker’s compensation or no-fault automobile insurance (Studdert and Brennan, 2001, Weiler, 1991; Abramson, 1990; Johnson, et al., 1989; AMA, 1987; Danzon, 1985b).

Siliciano (1991) has argued that malpractice law should be reformed so as to hold providers of care to the poor to a lower standard of negligence. Dewees et al. (1991) present evidence that suggests that increased medical malpractice frequency, severity, and probability in Canada are only partially explained by changes in the Canadian legal environment. As a result, they suggest that US type tort reforms are unlikely to have much of an impact on malpractice in Canada and recommend that a wider view of malpractice be taken where reduction in malpractice liability could be attained through amelioration of other factors which they claim have contributed more to the “crisis” (e.g., a general increase in the willingness to litigate, the impact that increased medical technology has on patients’ unrealistic perceptions of favorable outcomes, etc.).

Grady (1988) argues that the medical malpractice crisis has been caused, not by changes in the legal environment per se, but by increased technology. Grady argues that courts have responded by setting legal standards of care too high, especially with respect to what Grady terms “nondurable” precaution. Hence, in order to reform the malpractice system, one must change legal rules dealing with non-durable precaution where the standard of care is non-optimal rather than focusing on legal rules (such as the doctrine of informed consent and revision of the locality rule), which many commentators see as responsible for the malpractice crisis.

Since Grady’s work, a number of researchers have found that physician liability tends to react as predicted theoretically to changes in variables such...
as the specialty of the physician, the severity of the injury, etc. (e.g., Vidmar, 1995; Sloan, 1993; Taragin et al., 1992; Bovbjerg et al., 1989). Moreover, Farber and White (1994) found that plaintiffs who filed lawsuits mostly did so in order to gain information about potential negligence. In their sample, if evidence gathered subsequent to the suit indicated that a defendant was clearly non-negligent, then plaintiffs tended to drop the lawsuit.

To the extent that the increased physician productivity associated with such variables is associated with increased technology, these findings tend to support Grady’s contention that technological innovations lie at the heart of the historical increases in physician malpractice liability (see Olsen, 1997 for further evidence supporting Grady’s claim). However, as noted above and in contrast to Grady, these findings are supportive of the hypothesis that medical malpractice law is efficient. Obviously, if medical malpractice law is already working relatively efficiently, no need exists for medical malpractice reform.

Regardless of the widespread prescriptions for malpractice reform, actual public policy towards malpractice reform either passed by US state legislatures or found in US state and federal courts’ common law precedents has, in general, either explicitly rejected or have not even considered the proposed reforms discussed above. For example, courts have actually limited the ability of the parties to freely contract out of current medical malpractice liability rules (see, Tunkl v. Regents of the University of California, 383 P.2d 441, Cal. 1963). Only two states, Florida and Virginia, have replaced medical malpractice liability with a no-fault liability system and then only for infants who suffer neurological trauma during birth (Weiler, 1991; Epstein, 1988; Tedcastle and Dewar, 1988).

Rather than wholesale changes in the medical malpractice system, US courts and legislatures have, instead, focused on reforms whose apparent intent is to diminish the perceived crisis by reducing malpractice frequency, severity and probability. A few of the more popular US reforms, for example, include limiting plaintiff attorney fees, shortening the statute of limitations for malpractice claims, offsetting malpractice awards when plaintiffs have insurance which pays for injuries (collateral source rule), payment limits on malpractice awards, periodic payment of awards, arbitration, sometimes mandatory, of claims, and pre-trial screening of malpractice claims to establish merit (Olsen, 1996; Weiler, 1991; AMA, 1989; Bovbjerg, 1989; Danzon, 1985b). Malpractice reform of this nature has been extensive in the US. For example, by 1989 states had, on average, passed five of the ten most common reforms (Olsen, 1996). States continued to pass tort reforms affecting medical malpractice over the past decade as well. For example, the National Conference of State Legislatures (2007) notes this continued interest, with 36 states considering new reforms in 2006. Even though interest has periodically been expressed in malpractice reform at the federal level, to date reforms have been carried out at the state level (Thorpe, 2004; Waters et al., 2007).
Much of the empirical research in the literature attempts to measure the impact of such reforms on medical malpractice frequency, severity and probability. At the same time, such research also measures the impact that other legal, social or demographic variables have upon medical malpractice. Some of the earliest work assessing the impact of US malpractice reforms was done by Danzon (1984, 1985a, 1986). Using US state level data from the 1970s, she found that only a few of the malpractice reforms, dollar limitations on awards and the mandatory offset of collateral benefits (e.g., from health insurance), had a significant negative impact on medical malpractice severity (Danzon 1984). None of the other malpractice reforms that she tested had a significant impact on either frequency or severity. Further, she found no significant impact from any of the reforms upon frequency. Danzon did, however, find other variables that had a significant impact upon the frequency and severity of medical malpractice claims. For example, she found that urbanization was the most powerful explanatory variable. Further, pro-plaintiff laws passed prior to the 1970s were found to have a significant positive impact on both frequency and severity.

In contrast to Danzon, Sloan (1985) estimated the impact that various factors, including malpractice reforms, had upon malpractice insurance premiums (see also Viscusi and Born, 1995). Malpractice insurance premiums reflect, albeit imperfectly, increased malpractice frequency and severity through increased average benefits paid by insurers. Similar to Danzon, Sloan found that malpractice reforms generally had little impact upon malpractice insurance premiums. He found that only pre-trial screening panels had a significant and negative impact on premiums, while reforms requiring binding arbitration had a significant positive impact on premiums. In contrast to Danzon, who found that increased lawyer density had no significant impact on either frequency or severity, Sloan found that lawyer density had a significant positive impact on malpractice premiums. Danzon (1986) attributes this discrepancy between the two studies to Sloan’s failure to include urbanization as a control variable in his study. As Danzon (1984, 1986) notes, urbanization is likely to capture a number of relevant factors besides lawyer density which may also have a positive impact on malpractice frequency and severity and, eventually, upon malpractice insurance premiums as well.

Danzon (1986) updated her earlier study, which had only used data from the 1970s, incorporating data from 1975 to 1984. Danzon’s (1986) update of her earlier work, using an expanded data set to capture any lagged effect of the malpractice reforms, found both similarities and differences as compared with her previous empirical estimates. For example, similar to her earlier study, urbanization was positively related to both malpractice frequency and severity. However, urbanization was negatively related to the percentage of paid claims, which Danzon interprets to mean that urban areas have a higher than
average number of non-meritorious malpractice claims. Consistent with her
earlier study, lawyer density was found to have no impact upon either
malpractice frequency or severity.

In contrast to the earlier study, Danzon (1986) found several malpractice
reforms which tended to reduce malpractice frequency, including decreases in
the number of years before reaching the statute of limitations (i.e., a reduction
in the number of years during which claims can be filed) and mandatory collat-
eral source rules. In contrast, laws permitting voluntary binding arbitration had
the impact of increasing claim frequency. Similar to her earlier study, dollar
limitations on malpractice awards and the mandatory offset of collateral ben-
fits were found to reduce malpractice severity by approximately 23 percent for
the former and 11 to 18 percent for the latter. The allowance of voluntary bind-
ing arbitration was also found to reduce average severity by about 20 percent.
However, as noted above, since arbitration increased the frequency of
malpractice cases, these two effects tend to offset each other.

More recently, a number of studies have updated the literature on the
impact that reforms have upon malpractice liability especially as measured by
their impact upon the insurance market (e.g., Viscusi and Born, 2005; Danzon
et al., 2004; Born and Viscusi, 1998 and Barker, 1992). Using more recent
data that often extends into the 1990s and in a few cases into the 2000s, these
researchers have tended to confirm earlier findings that some malpractice
reforms do affect liability and by extension insurance premiums. As a general
rule, the more recent research has tended to show that caps on damages, espe-
cially on non-economic damages, have tended to reduce increases in malprac-
tice premiums. Viscusi and Born (2005) note that malpractice reforms tend to
perform as theory would predict, with a reduction in liability found to reduce
liability and insurance premiums. These researchers also tend to show an
improvement in insurance company profitability (Viscusi and Born, 2005 and
Barker, 1992).

The research consistently finds that caps on damages reduce liability and
insurance premiums. However, some researchers also find evidence that some
other types of malpractice reforms reduce liability and premiums. For exam-
ple, Viscusi and Born (2005) find evidence that some reforms that cap puni-
tive damages have an impact. Barker (1992) finds that reforms that codify the
standard of care decreased liability and risk, decreased premiums and
increased insurance company profitability. Kilgore et al. (2006) find that
statutes of limitation, which restrict the filing of malpractice claims to a
limited time period beginning with the discovery of an alleged negligent act,
did not tend to have a significant impact on premiums. However, in states that
passed the more restrictive statute of repose, which restricts filing to a fixed
time period regardless of discovery, premiums were found to be reduced
significantly. Danzon et al. (2004) find that limits on joint and several liabil-
ity that limit each defendant’s liability to only their share of the negligence also have the impact of lowering malpractice premiums.

Besides their impact upon malpractice frequency and severity, malpractice reforms have also been found to affect the decisions that plaintiffs and/or defendants face, once a claim has been filed. A malpractice claim can, ultimately, be dropped by the plaintiff, settled (which is a decision made by both parties), or continue to trial. In fact, in some states, the decision becomes more complex since voluntary binding arbitration is offered as a fourth possibility. To the extent that malpractice reforms affect these decisions, the reforms will have an indirect impact on malpractice, although not upon malpractice frequency (because the decision to file a claim occurs prior to the time when these decisions are made).

The earliest studies in this area, malpractice claim disposition, were also done by Danzon (Danzon, 1980; Danzon and Lillard, 1983). This later study (Danzon and Lillard, 1983) found that some malpractice reforms had a significant impact on the decisions to drop, settle or go to trial and, also, upon the dollar amount of verdicts for those cases going to trial. For example, they found that limitations on attorney’s fees caused plaintiffs to drop 5 percent more cases and decrease the proportion of cases litigated all the way to a verdict by 1.5 percent. Likewise, for settled cases, attorney fee limitations reduced the settlement size by 9 percent. For cases going to verdict, a mandatory collateral source offset rule reduced dollar awards by 15 percent. The impact of a number of reforms intended to reduce awards, such as dollar limits on the award and periodic payment of the award, were considered together and found to lead to a 30 percent reduction in average awards at verdict.

Hughes (1989), using data from the late 1970s, also corrected for potential selection bias and corrected the potential bias introduced by the inaccurate measurement of malpractice reforms used by earlier claim disposition studies. Correcting for these problems, he found that most of the malpractice reforms had no significant impact on the probability of settling a malpractice claim. Only limitations on lawyer fees and a collateral source offset rule significantly increased the probability of settling the claim. Hughes found that malpractice reforms had a much stronger impact on the decision to drop a malpractice claim. In fact, five of the eight reforms he considered, including limitations on awards and periodic payment, increased the probability that the plaintiff would drop the case.

Using more recent data from the 1980s in addition to data from the 1970s, Sloan et al. (1989) also estimated the effectiveness of the malpractice reforms in reducing malpractice liability primarily by their impact on probability and severity. Like other researchers, they also found that malpractice reforms had an inconsistent impact. For example, limitations on awards were found to significantly reduce severity, as found by earlier researchers. Likewise,
mandatory collateral source offset rules tended to reduce severity as well. Few of the rest of the reforms considered had a significant impact on severity. The only malpractice reform found to have a statistically significant impact on probability was the statute of limitations, where an increase in the number of years a plaintiff was allowed to file a claim increased the probability of payment.

Zuckerman et al. (1990) estimated the impact that tort reforms had upon medical malpractice insurance premiums, frequency and severity. Again, they found that malpractice reforms had an inconsistent impact on malpractice insurance premiums. The only reforms which significantly lowered premiums were those which limited the size of awards or those which limited the statute of limitations (Viscusi and Born, 1995, had similar findings with respect to malpractice premiums). None of the malpractice reforms were found to decrease frequency. Similar to other studies, Zuckerman et al. also found that limitations on awards, collateral source offset rules and limitations on attorney fees tended to reduce malpractice severity.

Synthesizing the empirical work estimating the impact of malpractice reforms, it becomes clear that not all reforms have a consistent impact on provider liability or insurance premiums. Reforms such as limitations on awards, collateral source rules, and shortened statute of limitations have had the most consistent impact on such liability. Most other reforms have only occasionally been found to have an impact on the medical malpractice liability of health care providers.

There are several public policy implications that may be gleaned from this empirical evidence. For example, if one takes the view that a crisis exists and needs amelioration, the empirical evidence highlights the types of reforms that would have the most probability of decreasing malpractice liability. Of course, as Hughes and Snyder (1989a, 1989b) illustrate, some care must be taken when making such public policy decisions. Care must be taken because the impact of the reforms is, theoretically, quite complex. For example, a reform may have the impact of decreasing severity, but changes in severity theoretically also have two competing impacts upon the frequency of claims. First, decreases in average awards give potential plaintiffs less incentive to file claims. However, decreases in average awards increase the incentive of potential defendants to be negligent, which, in turn, increases the number of potential claims. As a result, predicting the impact of a given reform upon malpractice liability, ex ante, is problematic.

3.4 Myths of Malpractice

As both the above discussion and the attached bibliography illustrate, there exists an extensive literature on medical malpractice from a number of differ-
ent perspectives. Previously, one of the deficiencies in this literature had been a failure to consider whether the malpractice system actually worked better than thought. The previous sections noted that neither malpractice cases nor large growth rates in malpractice liability are recent phenomena, even though many in the literature assume this to be the case. As a result, the possibility has always existed that the periodic medical malpractice crises are actually illusory. Most reforms, as outlined above, have little consistent impact on malpractice liability. One possible explanation for this is that the reforms, although intended to reduce liability, have not been well enough constructed to actually have that impact. Another possible explanation, though, is that the common law efficiently takes into account malpractice reform in the long run. That is, the law is flexible enough to overcome inefficient efforts to reduce malpractice liability.

An addition to the medical malpractice literature is an extended literature just begun when the earlier edition of the *Encyclopedia* was published a decade ago and summarized in a book by Baker (2005). This literature focuses on two main issues, both of which question the basis for believing that a medical malpractice crisis exists. The first investigates the crisis primarily from an empirical standpoint by questioning the basic premise that underlies both the malpractice crisis and the reforms that it has generated, that from the standpoint of society too many malpractice claims are filed. In contrast, this literature tends to show, as has actually been known for at least three decades, that in actuality too much malpractice occurs and there exist too few malpractice lawsuits (see also Sloan and Chepke, 2008).

The second issue investigates the reality of the common misconceptions that have typified the discussion in the political arena, the popular press, and the professional journals over much of the previous 40 years. This literature, discussed in more detail below, has shown a number of claims in support of the existence of a malpractice crisis to be false, including that the malpractice system worked differently for most of history and that juries often make decisions that favor plaintiffs and are not consistent with the evidence.

During each of the major malpractice crises identified above, major studies have been conducted of the existence of malpractice and its tendency to generate malpractice claims. The first study was done in the mid 1970s, commissioned by hospital and medical associations in California using only California data (see Mills, 1977). Somewhat surprisingly, at least at the time, the data from this study tended to show that the crisis in medical malpractice was in too little malpractice suits, not too many (see Mills, 1977; Baker, 2005 and Sloan and Chepke, 2008). The study found one out of twenty hospital patients injured as a result of medical care by doctors or hospitals. Only one out of six of these injuries was a result of medical negligence. Moreover, the rate of negligence to injury was found to rise as injury severity increased, with the
negligence rate reaching as high as 80 percent for the most severely injured patients. More recent studies have tended to confirm these results in different states and countries, including New York in the mid 1980s (Harvard Medical Practice Study, 1990), Utah and Colorado in the early 1990s (Thomas et al., 2002; Studdert and Brennan, 2000), Illinois (Andrews et al., 1997) and Australia (Wilson et al., 1995). As compared with the original California study, these studies have either comparable or higher levels of iatrogenic injury and instances of medical malpractice. More importantly, these studies also tend to find that the rate of malpractice claims is much lower than is the rate of medical negligence. In New York, for example, the Harvard Medical Practice Study (1990) estimated that approximately eight times as many patients suffer an injury caused by medical negligence as file a medical malpractice claim. Further, they estimated that 16 times as many patients suffer such an injury as the number who ultimately receive compensation through the tort liability system. The Harvard researchers note that these estimates understate the gap between negligent injury and compensation because the cases cited above include some filed where negligence did not occur. Baker (2005) notes that each of these studies came to comparable conclusions regarding the rate of negligent injury to the rate of malpractice lawsuits.

In addition to the misperception that there exist too many malpractice cases, rather than the reality that there exist too few, a number of other misperceptions exist about the functioning of the medical malpractice system. One of the most important misperceptions is that juries favor plaintiffs, especially in medical malpractice cases. Galanter (1990), Clermont and Eisenberg (1992) and Saks (1998) note that the prevailing public opinion among the public, lawyers, and public policy makers is that juries have a pro-plaintiff bias that makes it more likely that juries find in favor of plaintiffs and increases the damage award for successful suits. Related to the latter is the deep-pocket hypothesis that juries increase damage awards when the defendant has more financial resources (Vidmar 1993, 1995; Vidmar and Hans, 2007, Baker, 2005).

The evidence regarding plaintiff win rates is discussed above. Recall that plaintiff win rates remain relatively low, although some evidence exists that those rates rose somewhat from the 1960s to the 1980s. However, plaintiff win rates remain among the lowest in civil cases. For example, Daniels and Martin (1995) present evidence on plaintiff win rates from more than 20,000 jury verdicts from 16 states. For all civil trials in the data set that go to verdict, they find that plaintiff win rates are commonly over 50 percent and in some jurisdictions go as high as 70 percent. However, when the researchers use the same data set to consider only medical malpractice jury verdicts, they find that win
rates are much lower, usually below 50 percent and often below 30 percent. In fact, the overall win rate in all jurisdictions for medical malpractice jury verdicts is found to be only 30.4 percent.

A number of studies of all claims closed by medical malpractice insurers, including those closed by settlement and by jury trial, illustrate that plaintiffs tend to be paid only when the claim is valid. For example, Taragin et al. (1992) base their study upon the actual insurer classification of the claim as defensible, indefensible or unclear. When insurance companies classify the claim as indefensible they pay the claim 91 percent of the time; if classified as defensible they only pay in 21 percent of the cases, and if classified as unclear they pay in 59 percent of cases. Similar results are found in other closed claim studies. See Baker (2005) for a more detailed discussion of such studies.

Testing the “deep pockets” hypothesis, Vidmar (1995) notes that early evidence seemed to confirm the hypothesis. For example, Chin and Peterson (1985) found that defendants in jury verdicts from Cook County, Illinois between 1959 and 1979 tended to pay higher awards when the defendant was a health care provider or a corporation as opposed to an individual. For more severe injuries, health care providers paid as much as four times that of defendants who were individuals. Other studies seemed to confirm these results (see Vidmar, 1995; Peterson, 1987; Daniels and Martin, 1986; Hammitt et al., 1985).

A number of researchers noted problems with these results, including the differences in the types of cases likely to be seen at trial (e.g., Daniels and Martin, 1986; Bovbjerg et al., 1991; Danzon 1985b). Vidmar (1995) notes that a number of other studies have concluded that the best predictor of the size of the award is not the financial resources of the plaintiff, but, rather, the severity of the injury and other variables that would be expected to affect damages such as physician specialty, etc. (see also Sloan, 1993; Taragin et al., 1992; Bovbjerg et al., 1989; Vidmar and Hans, 2007).

Vidmar (1995) directly tested the deep pocket hypothesis in a series of jury experiments using real jurors presented with hypothetical cases and found no statistically significant support for the deep pocket hypothesis. Vidmar (1995) also used jury experiments to test the hypothesis that jurors decide cases differently than legal professionals such as judges, and again, found no statistically significant support for the hypothesis. These studies, and others, suggest that jurors tend to do a reasonable job in deciding cases, especially as compared to lawyers.

One final issue to be discussed is that of “defensive medicine”. Defensive medicine is the hypothesis that fear of a lawsuit leads doctors to order unnecessary tests or procedures. The issue of defensive medicine is complicated by the fact that tort law is intended to give potential defendants an incentive to take appropriate caution in avoiding accidents. Hence, in testing whether or
not malpractice law causes defensive medicine, it is not sufficient to show that negligence changes doctors’ behavior, but it must also be shown that the change was not cost effective. Baker (2005) notes that most studies of defensive medicine do not make this distinction and instead simply focus on the simpler issue of whether negligence causes doctors to change their practice patterns.

A great deal of attention, however, has been devoted to this issue, with differing conclusions from different studies. Early studies relied upon opinion surveys of doctors, either simply asking doctors retrospectively whether they practice defensive medicine or giving them a specific clinical case and asking questions about how they would treat the patient. Baker (2005) discusses these studies in detail and notes especially for the clinical surveys that, although generally discredited, they tend to demonstrate that doctors say they practice defensive medicine less often than commonly thought.

Studies do exist that find evidence of defensive medicine, as defined above. For example, Kessler and McClellan (1996) used Medicare (US government provision of medical insurance for the elderly) records to estimate the prevalence of defensive medicine. They focused only on the treatment of heart disease and used the Medicare data to identify systematic differences in treatment patterns across states with different types of tort reforms. They found that treatment costs for heart disease grew significantly faster in states without those reforms more likely to reduce physician liability. The size of the impact was dependent upon when the state level tort reforms were enacted, with longer time periods under tort reform having the expected larger impact.

Kessler and McClellan (1996) also found that the Medicare data indicated that lower costs found in reform states did not significantly change health outcomes. In other words, they found evidence of defensive medicine. Dubay et al. (1999 and 2001) also found evidence of defensive medicine in their studies of cesarean section rates and prenatal care. In these cases, however, although malpractice fears as measured by malpractice insurance premiums did impact practice patterns, the impact on health outcomes was either not present (prenatal care) or very small (cesarean sections). Baicker and Chandra (2004) have attempted to extend Kessler and McClellan’s (1996) techniques to a number of other types of illnesses and found little evidence of defensive medicine. Thus, the available evidence suggests that if defensive medicine does occur its impact on health care costs is probably relatively small.

3.5 Conclusions

The regulation of medical professions occurs through two main methods. One method is the ex ante regulation occurring with the licensing of medical professionals, ostensibly to assure quality. However, licensing regulation is
shown above to have an inconsistent impact on quality and to often be used as an entry barrier to increase prices and profits of the professionals being regulated. Second, the quality of care offered by medical professions is also regulated ex post using medical malpractice lawsuits.

Theoretically, health care providers are found liable for negligence under medical malpractice law in order to optimally deter them from negligent treatment. Although the literature, as noted above, consistently assumes that medical malpractice fails to meet the goal of optimal deterrence because of imperfections in the legal system, it also consistently notes that such deterrence is a valid and necessary public policy goal. More than any other event, increased health care provider liability over the past three decades has resulted in the common perception that the medical malpractice system is both imperfect and in need of reform. Some care, however, must be taken in this conclusion as recent evidence indicates that, contrary to the common perception, substantial growth rates in health care provider liability for malpractice is not a new phenomenon nor limited to the US. Moreover, substantial evidence exists that increases in physician liability occur primarily because of an arguably efficient response to changes in variables, such as increases in the severity of the injury or increases in the medical cost of treating injuries.

Much of the early literature on medical malpractice has focused on either the need for, and the most efficient type of, malpractice reform or the impact that existing reforms have had on medical malpractice. In the past decade, research has begun to test the assumptions upon which the need for malpractice reform is based, including questioning whether too much rather than too little malpractice exists, whether juries behave as poorly as assumed in some of the literature, and whether defensive medicine is prevalent. Given the questionable basis for the common conclusion that malpractice reform is necessary, that a smoothly working malpractice system has suddenly gone into crisis and recent evidence that tends not to show a need for reform, the possible efficiency of medical malpractice, at least on a second-best level, must be seriously evaluated by researchers.

Common medical malpractice reforms in the US have had an inconsistent effect on medical malpractice liability. Some reforms, such as limitations on awards, collateral sources rules, and shortening the statute of limitations, have been found to have a fairly consistent impact on malpractice liability. However, other reforms have been found to have either no significant impact or an inconsistent impact on malpractice liability. As noted above, the inconsistent impact of reforms intended to reduce malpractice liability may mean, as some researchers have suggested, that more care should be paid to the nature of malpractice reform. However, it may also be the case that the reforms are ineffective because the law (e.g., judges and juries) is flexible enough to obviate the reforms and still impose increased liability on health care providers.
Research mostly done in the past decade has begun to highlight some of these issues and has tended to show that the medical malpractice system works much better than is often claimed. Even though some researchers and many public policy makers continue to make claims to the contrary, juries tend to work relatively well in allocating negligence and in deciding on appropriate level of awards. Although some evidence of defensive medicine, the wasteful increase in tests or other treatment to avoid liability does exist, the evidence tends to find that the impact of such defensive medicine is relatively small and for some diseases or treatments not present at all. In fact, evidence seems to suggest that the most recent malpractice crisis is tied more to the failings of the insurance market than it is to failings in the malpractice system itself. Taken together, the literature and, especially, the most recent results tend to suggest that the malpractice systems are working fairly efficiently.

BIBLIOGRAPHY


Dewees, Donald N., Michael J. Trebilcock and Peter C. Coyle (1991), ‘The Medical Malpractice


Lindsay, Cotton M. (1973), ‘Real Returns to Medical Education’, *Journal of Human Resources*, 8, 331–48.


Tunkl v. Regents of the University of California, 383 P.2d 441, Cal. 1963.


