Deterrence and Liability for Medical Negligence: Theory and Evidence

Paul Fenn
Nottingham University Business School

Alastair Gray
Health Economics Research Unit
University of Oxford

Neil Rickman
Department of Economics
University of Surrey

Robert Young
Nottingham University Business School

Paper to be presented at the 19th Annual Conference of the European Association of Law and Economics, Athens, 19-21 September 2002.

Abstract

Patient compensation systems are means by which two objectives can be pursued: the cost of the harm can be transferred away from the patient (the "compensation" objective); and the doctor can be given an incentive to take appropriate care to avoid making mistakes which may harm their patients (the "deterrence" objective). All patient compensation systems attempt to deliver these objectives jointly with varying degrees of success, and with varying administration costs. The paper begins by reviewing what theory and evidence exists from various countries in relation to the impact of both existing and proposed alternatives in delivering the deterrence objective. Additionally, an empirical study was undertaken to explore for the first time inter-hospital differences in recent clinical negligence experience under the current UK system, which is based on enterprise fault liability. Data were collected from hospitals in relation to the numbers of open, closed and outstanding claims in recent years. Information about insurance arrangements, including subscriptions, voluntary excess levels and risk management discounts have also been collected. These data are analysed in the paper to identify those hospitals with claim rates significantly higher than would be expected given their casemix, and to discover whether these hospitals are systematically different in relation to their exposure to litigation risk as measured by their voluntary excess levels. The paper concludes with some policy implications.

1. Introduction

Medical treatments are inherently risky. There are occasions when patients are harmed as a consequence of their treatment or absence of treatment. Patient compensation systems are a means by which two objectives can be pursued: the cost of the harm can be transferred away from the patient (the "compensation" objective); and the doctor can be given an incentive to take appropriate care to avoid making mistakes which may harm their patients (the "deterrence" objective). All patient compensation systems attempt to deliver these objectives jointly with varying degrees of success, and with varying administration costs.

The two key design features of any patient compensation system are:

1. Eligibility: what must the patient prove in order to receive benefits?
2. Responsibility: who is immediately responsible for paying the benefits from a successful claim by a patient?

Table 1 presents a matrix that seeks to classify existing patient compensation systems in relation to these features. Along the top it distinguishes (as we move from left to right) between fault-based and non-fault-based compensation schemes. Moving from top to bottom, it distinguishes the identity of the body responsible for paying compensation if this is required.[1]

[1] The matrix ignores the role of insurance. Where doctors or hospitals are liable for compensating patients, they can transfer this risk through liability insurance or some other pooling mechanism. In that case, the impact of the compensation system on incentives will depend on the extent to which some financial pressure is brought to bear on the immediate risk-bearers (perhaps through excesses or experience-related premiums).

Table 1: Possible arrangements for compensation schemes

	Basis for eligibility		
Immediate risk-bearer	*Fault and cause and severity*	*Cause and severity*	*Severity*
Doctor/clinician	Individual fault liability (e.g. pre-1990 tort in UK NHS)		
Hospital	Enterprise fault liability (e.g. post-1990 tort in UK NHS)	Strict enterprise liability (e.g. Workers'Compensation in US)	
Government		Scandinavian no-fault	Social security

Some of these schemes may do better than others at securing compensation at low administrative cost. Some may do better at sharpening the incentives of health care providers to take care.[2] This paper attempts to discover what evidence exists in relation to the impact of existing patient compensation systems in delivering the second of these objectives.

Before seeking to examine the evidence in more detail, we address some conceptual issues relating to the economic role of negligence and causation in the area of compensation for medical injury. The paper then proceeds to review the empirical evidence on deterrence and defensive medicine from other countries before focussing

[2] In some systems, the two objectives have been explicitly separated; one mechanism can be set up to deliver compensation to patients, another to secure medical accountability.

on the current situation in the UK. The final substantive section presents some results from an exploration of inter-hospital variations in claims using UK data.

2. Conceptual issues

It is well known that beneficial activities can lead to costly outcomes ('harm', 'injury'), e.g. on the road or in the operating theatre. Such costly outcomes can be reduced if the beneficial activities themselves are cut down, or if those involved take care to avoid them. However, to the extent that care is also costly, people may need to be given incentives to provide it. One natural incentive is to make the person causing the harm liable for the costs involved, if s/he fails to supply care beyond a sufficient threshold (i.e. behaves 'negligently'). This potential attribution of fault provides a deterrent against insufficient care levels. In theory, the appropriate level of care should maximise the net gains to society from the beneficial activities involved: this means that the marginal social benefit from an extra unit of care should equal its marginal social cost. In other words, the extra benefit to society (in terms of reduction in injury rates and their associated costs) from an extra unit of care should just equal the extra resource cost to society of the extra unit of care itself. It should be understood that the notions of benefit and cost here are broader than simple monetary amounts.

In a world where the standard of care is known to everyone, and observable (to individuals and courts), it is straightforward to show that negligence liability would produce socially optimal levels of care (i.e. deterrence). However, common sense and the observation that many people buy liability insurance, tell us that these conditions are strict and that, in practice, overwhelming information problems prevent such a

result. For example, courts cannot determine precisely what care a clinician has taken, while it is often the case that opinions differ as to what constitutes an appropriate standard of care. This can have numerous implications for the successful operation of a negligence system. Particularly important here is that if the care-threshold and courts' abilities to apply it are unpredictable, parties worried about mistakenly being found liable may over-invest in care. In the medical context, this is what is known as 'defensive medicine'.[3] Thus, when assessing a negligence rule for medical negligence we should, ideally, be able to compare the deterrence benefits of such a rule, with the defensive costs that it may induce in practice. Danzon (2000a, b) observes that negligence-based liability for medical malpractice appears to involve higher administrative costs than alternative means of providing compensation (see Section 3.2 below). Therefore, a necessary condition for the adoption of negligence-based compensation is that it provides net deterrence benefits.

One, apparent, way to overcome this problem is to move to a system of 'strict liability', where proof of causation is sufficient to trigger compensation from the party whose actions led to the harm. Because this payment is made regardless of the injurer's level of care, it overcomes problems associated with unpredictable care thresholds. Also, to the extent that the injurer faces the costs of his/her actions, it can be shown to provide optimal incentives for care (i.e. deterrence). Once more, however, real-world imperfections make this unlikely in practice. Now, the question surrounds whether courts can observe damage levels and set suitable penalties

[3] Danzon (2000b) notes another (of many) problems in the case of medical negligence. If the care threshold is set by practitioners' reports of 'best practice' (as is typically the case), this may lead to sub-optimal negligence rules to the extent that systematic errors exist within this 'best practice'.

accordingly. Again, the prospect of excessive damages can stimulate excessive (i.e. defensive) levels of care and we must again recognise the need to trade-off deterrence benefits with defensive costs when evaluating such schemes.[4]

Another important consideration when evaluating the deterrence effects of both rules above is the role of third-party payers such as insurers. To the extent that it is common for potential tortfeasors to shift risk onto insurers they may avoid facing sufficient costs to induce appropriate care. A range of measures can mitigate these concerns, including experience-rated premiums and co-payments/deductibles.[5,6] An alternative third-party payer may be the individual's employer (e.g. a clinician's hospital).[7] Under such 'enterprise liability schemes' (a version of which has existed in the UK NHS since 1990; see also Weiler, 1991; Morlock and Malitz, 1991), the hospital may seek to shift risk onto an insurer and issues similar to those above arise. Of course, in order to ensure that employees have incentives to supply care, the enterprise will need to have mechanisms in place to monitor, record, investigate and, possibly, punish any acts for which it is held liable. Even if this is possible, it will entail some measure of cost, so justification of such schemes requires that benefits exist to offset this. These may include the hospital being the appropriate risk-bearer

[4] Cummins *et al.* (2001) show that it is not clear *a priori* whether negligence or strict liability generates higher care levels. For negligence to produce this, a sufficient condition is that courts are 'sufficiently accurate' in assigning fault.

[5] Co-payments and deductibles are the portion of the claim that the insured must meet when claiming on his/her own policy. As in other areas of insurance (e.g. automobile), they are aimed at providing an incentive against excessive claiming and, therefore, in favour of taking care.

[6] The insurer may also seek to monitor its policy-holders' behaviour but the transactions costs of doing this may be prohibitive or damaging to the efficiency of the underlying liability rule.

[7] In the case of US health plans, the hospital and insurer jointly bear the risk.

for systemic risks (as opposed to case-level ones): it designs the risk management systems and is responsible for their operation.

Cummins *et al.* (2001) identify strict liability schemes with 'no-fault' set-ups: in both cases payment of compensation is not contingent on the attribution of fault. Automobile accident compensation in Canada, New Zealand, Australia and some US states would be examples of individual policies here, while Workers' Compensation in the US would illustrate the enterprise-level equivalents. However, the class of no-fault policies is broader than this. In the medical negligence context, for example, the Swedish Patient Compensation Insurance (PCI) provides compensation (if cause is proved) from government revenue. Deterrence is achieved, in principle, by a Medical Review Board to investigate and reprimand clinicians found to be at fault for the injuries in question. The effectiveness of such 'decoupling' of fault and compensation is an important empirical question. Another version of no-fault schemes can be found in New Zealand's Accident Compensation Scheme (ACS). Here, originally, even causation was not a significant issue in triggering compensation. As we shall see, however, instructive changes have taken place since the scheme's inception in 1972.

Apart from any theoretical arguments surrounding the relative merits of no-fault schemes, Danzon (2000a, b) provides some interesting comments. To the extent that deterrence is not an aim of these schemes, she raises the question of whether compensation for medical mishaps is an appropriate concept. In particular, why should someone be compensated for illness when caused by medical care but not if the same was the result of birth defects of genetic inheritance? Perhaps social

insurance programmes are the most equitable way to provide 'deterrence-neutral' compensation.

3. Evidence on deterrence and defensive medicine

A number of issues arise when comparing fault-based with no-fault compensation systems. The relative cost of running the two types of system and the compensation available in each case are dealt with in Section 3.2 below. In this section, we restrict attention to the *incentive* effects that the two systems create. These can be divided into two broad categories:

- *Deterrence effects*: An important objective for fault-based schemes is to provide practitioners with incentives to take care and, therefore, to restrict the potential for injuries. As we have seen, this is achieved by seeking to identify those responsible for these injuries and requiring them to face the costs of their actions. Of course, as Cummins *et al.* (2001) observe, it is possible for no-fault schemes to perform a similar job to the extent that they penalise the culprits whose actions have been found to have *caused* the injuries. Alternatively, it is possible to de-couple the deterrence and compensation roles of a particular system by reporting the results of no-fault investigations to regulatory bodies (or employers) to take separate actions against the culprits: Studdert and Brennan (2001) advocate such measures (as noted in Section 3.2). As Danzon (2000a, b) makes clear, it is the deterrence effects of no-fault that must be its main justification, which accounts for our interest in this issue here. It is thus important to know whether fault-based systems achieve deterrence and how much they achieve relative to no-fault

based ones: movement towards the latter may generate unseen costs in terms of additional injuries.

- *Defensive medicine*: There is (in theory) an 'optimal' level of deterrence – that which equates the marginal costs and benefits of taking an extra unit of care. However, as we have seen, in situations where the negligence standard is unclear, or where practitioners are excessively risk averse, they may over-react to deterrence incentives and supply 'too much' care. Such 'defensive' behaviour increases the cost of the systems in which it takes place.

When looking for evidence of these phenomena one needs to bear in mind that empirical work faces considerable difficulties relating to what would be optimal behaviour in a given setting. For example, if we observe more care being taken under one system than another, we can only define this as sufficient, insufficient or defensive (i.e. excessive) deterrence by referring to an ideal level of care – something we cannot readily observe. Nevertheless, some interesting studies have been undertaken in this area and we summarise some of the key findings below.

Table 2 summarises the results of several key papers seeking to measure the relative effects of different compensation schemes on care levels. The papers are divided according to whether they describe themselves as relating to the "deterrence" or "defensive" effects of compensation although, as pointed out above, it may be hard to distinguish between these in practice. We begin with the papers on deterrence. Two broad approaches to analysing care levels can be discerned here. The first involves the use of data sets and here the key requirement is that there is variation in the types of compensation scheme being used. This may be variation across a country at

a given point in time ('cross-sectional' studies) – as in the variation in schemes across American States – or across time ('intertemporal' studies) – perhaps as a result of reforms leading to a change in regime. The second approach involves 'creating' one's own data by interviewing physicians (etc.) about their care decisions and about their perceptions of the extent to which they are deterred by tort.

Table 2: Summary of results on deterrence and defensive effects

Source	**Context**	**Deterrence/defensive effect of tort?**
Papers on deterrence		
Cross-section and intertemporal studies		
Landes (1982)	US Auto	Yes
Kochanowski and Young (1985)	US Auto	No
Zador and Lund (1986)	US Auto	No
McEwin (1989)	Australia, NZ Auto	Yes
Devlin (1992)	Quebec Auto	Yes
Cummins *et al.* (2001)	US Auto (all states)	Yes
Interview-based studies		
Harvard Study (1990)	US Medical (NY)	No (occasional Yes)
Weiler *et al.* (1993)	US Medical (NY)	Yes (insignificant)
Papers on defensive medicine		
Claims history studies		
Reynolds *et al.* (1987)	US physicians	Yes
Localio *et al.* (1993)	31 acute hospitals, NY	Yes
Baldwin *et al.* (1995)	Washington State practices	No
Sloan *et al.* (1997[1])	31 Florida counties	No (occasional Yes)
Intertemporal studies		
Kessler and McClennan, (1996)	US heart patients, 84, 87, 90	Yes
Kessler and McClennan, (2000a)	US heart patients, 84-94	Yes (diminished)
Kessler and McClennan, (2000b)	US heart patients, 84-94	Yes (theraputic < diagnostic)

Consider first the data-based cross-sectional/intertemporal studies. These are drawn from countries where automobile compensation systems vary (or have varied over time) between fault-based and no-fault schemes: the US, Australia, New Zealand and Canada being examples. In each case, the studies seek to determine how the number of fatal automobile accidents is affected by the type of system. Early studies in the US (Landes, 1982), Australia and New Zealand (McEwin, 1989) and Canada (Devlin (1990) all discover that fatal accident rates are significantly higher under no-fault regimes: for example, Landes finds an increase in these rates of between 2–14%; McEwin finds an increase of 16% and Devlin finds an increase of 9%. The studies attribute these increases to the fact that drivers under the no-fault systems take less care than under the fault-based ones: a deterrence effect. However, other studies in the US, using later data (Kochanowski and Young, 1985; Zador and Lund, 1986) fail to corroborate these results: both studies find no significant difference in accident rates across states with different compensation systems.

In a recent and comprehensive study, Cummins *et al.* (2001) seek to explain this difference in results. They suggest that a problem with previous studies is that they do not take account of the two-way association that may exist between the accident rates and choice of compensation system. Thus, the above studies assume that the type of system 'leads to' the accident rate, but it is possible that a reverse link is also present: states with high accident rates may choose to use no-fault systems (perhaps because of their administrative simplicity in dealing with a high volume of cases). Ignoring this effect may bias results. Cummins *et al.* present evidence consistent with the presence of such endogeneity. They find no relationship between the compensation system and accident rates before controlling for endogeneity. However,

once it has been controlled for, no-fault accident rates are between 5.5% and 7.8% higher than under tort. This appears to explain earlier studies' conflicting results, and to indicate the presence of a deterrence effect due to tort.

We next consider interview-based studies. The Harvard Study of medical malpractice in New York (1990) contains a selection of interviews with physicians (Ch. 9). These interviews ask about attitudes towards New York's tort-based malpractice system and presented a set of vignettes asking physicians to explain how they would handle some hypothetical cases. The study is interesting because it provides an attempt to look at deterrence in a medical context. However, we may feel ambivalent about the methodology adopted: if tort achieves deterrence by forcing physicians to take (costly) care, they may feel some degree of antagonism towards the system.

Bearing this observation in mind, the Harvard Study concluded that "physicians are quite likely to downplay the specific deterrent effect of malpractice litigation", categorising such litigation as an "irritating nuisance rather than something that affects the way they practice medicine." (p. 9-67). In particular, physicians believed that the punishment meted out for negligence was related to whether a trial (and press coverage) was involved, rather than the extent of the injury. They also believed that this punishment was a long way removed from their actions, in terms of the delays involved in the tort process. Thus, to the extent that the appropriateness and timeliness of punishment for negligence are important to the deterrence effect (as they are), physicians implied a negligible deterrence effect due to tort. The authors

note an interesting caveat, however: the physicians typically over-estimated the risk of being sued. Perhaps they supply extra care without realising they are doing so.

In subsequent work, Weiler *et al.* (1993) perform statistical analysis with these data. Their results suggest that tort liability reduces the number of negligent injuries per admission by 29%, but this effect is statistically insignificant. They note however that, given the problems of work in this area and the above evidence of over-estimating risk, their results may well be consistent with liability rules having a deterrent effect in medical cases.

We now turn to the studies in Table 2 that seek to examine the question of defensive medicine. Again, two approaches to this question are contained in the work we survey. The first attempts to link physician practice decisions with their (or their employer's, or their region's) previous claims history – a proxy for the risk of tort liability. The idea here is that some procedures are "safer" than others and might, therefore be employed defensively. The second type of work seeks to compare care choices over time, looking at how they have responded to changes in the liability rule (intertemporal studies).

Taking claims history work first, early American studies by Reynolds *et al.* (1987), Localio *et al.* (1993) and Baldwin *et al.* (1995) provide conflicting results. Reynolds *et al.* use an interview approach to question physicians about specific changes made to their practices in response to claims risk: for example, changes to record keeping, tests, treatments, follow-up visits. When added to liability premiums, they estimate the additional cost of this care at 14.1% of gross practice revenue (in 1984).

The study by Localio *et al.* (1993) tries to link the incidence of cesarian deliveries in New York State in 1984 to physician malpractice premiums, on the grounds that cesarians are a lower risk method of delivery than vaginal birth. Having controlled for the clinical risk of cesarian delivery, patient socioeconomic status and physician/hospital characteristics, they find that the odds of a cesarian are three times more likely in high premium areas. They interpret this as evidence of defensive practice.[8] Baldwin *et al.* (1995) perform similar analysis for 1988/89 data from Washington State; they include antenatal care (e.g. obstetric ultrasound use and referral/consultation) along with the incidence of cesarian delivery as measures of care that obstetricians can supply. In contrast to Localio *et al.*, these authors find no link between claims history and these 'defensive' variables, suggesting that tort does not encourage defensive care.

Subsequent work by Sloan *et al.* (1997[1]) argues (as did Cummins *et al.* earlier) that endogeneity may explain these conflicting results: whilst claims history may influence care choice, so a physician's treatment preferences over time may have influenced his/her claims history. Having controlled for such effects, Sloan *et al.* fail to find substantial evidence of any link between clinicians' previous claims history and (i) the choice between cesarian and vaginal birth; (ii) the use of antenatal testing, and (iii) mothers' satisfaction with their treatment levels. In general, they find no

[8] One could argue that this result may simple imply more deterrence as a result of claims history. However, to the extent that the theoretically optimal level of care is not expected to vary with physicians' claims history (i.e. with *private* marginal benefits from care), Localio *et al*'s interpretation of their result may be legitimate.

significant differences as the claims history changes (with the exception of antenatal tests but, here, the signs are not always in the direction of defensive care levels).

Now consider the intertemporal work in the three papers by Kessler and McClennan (1996, 2000a, 2000b). These note that tort liability reform throughout the US in the '80s and '90s should have influenced treatment levels because of associated changes in physicians' exposure to litigation risk. Using data on heart treatments in 1984, 1987 and 1990, they estimate that reforms aimed at relaxing tort (e.g. damage caps, contingency fee caps, shifts to no-fault) decreased expenditure on heart treatments by between 5% and 9% over the period. Given that the authors control for the outcomes of treatments, this can be interpreted as evidence of defensive medicine under tort: the 'same' cases, with the 'same' results, received less treatment expenditure under tort.

Once again, a danger of endogeneity is present, this time because of US managed care. Managed care programmes have been shown to reduce health care expenditure so, if they also seek to adopt less strict tort regimes (perhaps to control liability premium costs), a negative relationship between 'relaxed' tort and treatment expenditure could be due to the presence of managed care in states with tort reform. Taking this into account, Kessler and McClennan (2000a) find a reduced, but still significant, defensive effect: now tort reform reduces treatment expenditure to between 4% and 5%. In Kessler and McClennan (2000b), the authors seek to identify which areas of practice the defensive care they have identified are located. Their results suggest that larger defensive effects occur in diagnostic care than therapeutic care.

These studies indicate the considerations (in terms of data and method) necessary for examining the questions of care levels under compensation systems. Given the difficulties of performing such work, it is perhaps unsurprising that a mixture of results is available: perhaps one should adopt Weiler *et al.*'s (1993) view that, because statistical results will inevitably be hard to find, some generosity is needed in interpreting those available.

If we take the most recent of the studies (Cummins *et al.* and Kessler and McClennan's work) there seems to be reasonable evidence that fault-based regimes generate more care than no-fault alternatives. Whether this care level is excessive is hard to judge (for reasons set out earlier) but, to the extent that it is, the appropriate solution may not be abandonment of the fault option. Instead, clearer negligence standards, data on claims experience and efficient procedures for adjudicating on claims may be sufficient to generate appropriate levels of deterrence (e.g. see Farber and White, 1991; Hughes and Savoca, 1997).

4. The UK context

The UK government have recently announced plans for a White Paper on patient compensation, raising the possibility that the current tort system may be reformed to a greater or lesser extent. However, what has sometimes been overlooked has been the impact of gradual structural change over the last decade on the way in which the current tort system operates in the UK. Since 1990, the health service has been decentralised to a significant degree such that individual hospitals have acquired considerable financial autonomy and have adopted commercial accounting practices.

Over the same period, moreover, the responsibility for compensating injured patients has, almost unnoticed, shifted first from the individual clinician to the hospital, and now finally to the NHS Litigation Authority (NHSLA) as the central agency set up to pool litigation risks through what is known as the Clinical Negligence Scheme for Trusts (CNST). The NHSLA has, from April 2002, taken financial responsibility for 100% of all claims against NHS hospitals. Prior to this date, under the terms of the CNST, hospitals had to retain part of the cost through choosing an "excess" level, below which they were responsible for the patient's claim. The decentralisation of accounting responsibilities for small value claims placed an additional burden on hospital management, and led to difficulties in producing consolidated estimates for the NHS accounts. These difficulties were behind the move to shift all financial responsibility for claims to the NHSLA, a move which should markedly improve future public information about the frequency and cost of clinical negligence in England. Now that the NHSLA is responsible for all claims, it should be in a position to report on national trends in the frequency and cost of medical litigation, as well as identifying those activities and procedures most at risk of litigation. In principle, data on claims could be coordinated with data on adverse events as reported to the National Patient Safety Agency (NPSA).

These potential benefits have materialised as a consequence of the transfer of responsibility for claims from hospitals to the NHSLA. However, as discussed in section 2 above, it is usually recognised that those who cause injuries should themselves face at least some of the injury costs, in order to provide potential injurers with an incentive to take care. In the health care sector, this issue is complicated by the fact that patients may be injured due to the interaction of

multiple factors leading to organisational, rather than individual, failures. In those circumstances, it becomes important to provide hospital managers with incentives to take responsibility for identifying system failures and implementing risk management procedures. Arguably, the combined effect of switching financial responsibility for negligence from individual clinicians to hospitals, and imposing a minimum excess level as a condition of pooling risks through the CNST, represented a coherent policy in this respect in the UK during the 1990s. Although hospitals could pass on the cost of their below-excess claims to health care purchasers, this in itself provided some kind of financial discipline. By now reducing excess levels to zero, the remaining financial incentives to pursue good risk management practices are through CNST subscription discounts.

One such discount is given by the NHSLA to hospitals who achieve certain assessed risk management standards. While these standards are designed to include the presence of, *inter alia*, adequate incident reporting and complaints management systems, they are a reflection of processes, not outcomes. A second discount which does potentially give hospitals a financial stake in reducing the number and cost of claims is given by the NHSLA in relation to hospitals' claims experience. However, it is not particularly clear how claims experience is measured for this purpose. Newly opened claims may turn out to be unjustified, or have low settlement values. Claims closed with a known payment may reflect risk management decisions taken decades prior to the year of settlement. For some hospitals, small enough to experience low absolute numbers of claims, this information would in any case be thin, and sufficiently variable to mis-represent their relative risk in most years. In any case,

unless these discounts are made more transparent, they may not succeed in providing the signals they are designed to send.

5. Some evidence from the UK

Data

None of the studies reported in section 3 above involved UK data. This, of course, reflects the lack of variation (across the country or over time) in the liability rules implied by the compensation system – as well as the relative lack of data on claims. In this paper we attempt to fill this gap in the literature by confronting each of these problems, and at the same time shed some light on the deterrence effects of enterprise liability in the UK.

The lack of variation in liability rules is, of course, insurmountable: all hospitals in the UK face the same system of civil law, and the basis of liability has remained unchanged for centuries. However, as pointed out in the previous section, it is not true that all hospitals faced the same expected cost of litigation. Each hospital until recently has chosen an excess level under the pooling scheme (the CNST), and this determines the subsequent exposure to liability risk. The payment for CNST cover varies depending on the excess level chosen, and, to a limited degree, on the risk management standards applied and the claims experience observed over previous years. Because of the limited nature of experience rating, hospitals with low excess levels face a lower expected cost from increased litigation than those with high excess levels. Consequently we can in principle test whether this variation in liability risk has an impact on hospital care levels. In this study we use the observed changes in claim frequency and claim costs to measure the impact of changes in care levels.

Table 1 shows the variation in CNST excess levels and risk management discounts across all English member hospitals:

Table 1 here

Claim frequency can be measured in a number of ways: for a sample of English hospitals we obtained data on the number of opened claims, the number of paid claims, and the number of claims currently outstanding. Figure 1 below shows the distribution of the number of new claims in 2001 for those hospitals that responded to our survey:

Figure 1 here

It is clear from Figure 1 that, for measures of claim frequency, there are substantial numbers of hospitals with zero observations. This suggests that count data methods will be appropriate for analysis of these data, and this is the approach we take below when analysing the variations in claim frequency across hospitals.

Claim severity – average claim costs – can also be measured in different ways. We have data on the breakdown of claim costs into the award paid to the claimant, the claimant's legal costs, and the defendant's legal costs. The descriptive statistics for claim severity are shown in Table 2 below.

Table 2 here

As far as other factors potentially influencing claim frequency and severity are concerned, we consider the possibility that the most important of these relate to the size and type of the hospital. Clearly, the number of admissions or treatment episodes at a particular hospital will be a factor determining the number of claims; moreover, it is possible that larger hospitals, with more admissions, will have a smaller or larger claim *rate*, depending on whether risk management activities are subject to increasing or decreasing economies of scale. We explain below how this hypothesis can be tested within a count data approach to estimation. Secondly, the nature of the admissions or treatment episodes will presumably influence the frequency and cost of claims: maternity hospitals, and those with a large proportion of acute beds, may be more open to litigation than others, for instance. Table 3 summarises the data we have in relation to hospital size and type.

Table 3 here

Methodology

For a given hospital, the process by which observed data on the number of clinical negligence claims are generated over time can be characterised as a Poisson process with a constant rate of occurrence, μ. The observed number of claims will clearly depend on the population at risk – in this case the number of finished consultant episodes (FCEs) in a given year, N. Consequently, the expected number of claims occurring in a given year at a given hospital is the product $N\pi$, where π represents the mean probability of a finished consultant episode resulting in a claim. Given the assumed Poisson process, this implies that the observed number of claims (y) in a hospital in a given year is distributed with density

$$f(y; N, \pi) = \frac{e^{-N\pi}(N\pi)^y}{y!} \tag{1}$$

While we can observe N for each hospital, the parameter π is a latent variable, which can nevertheless be modelled as a function of observed covariates and unobserved random variables. With a conventional loglinear specification of this function, we have

$$\pi = \pi(\mathbf{x}, \boldsymbol{\beta}) = e^{\mathbf{x}'\boldsymbol{\beta}+\varepsilon} \tag{2}$$

where $\mathbf{x}$ represents a vector of observed covariates which are assumed to influence the risk of litigation, and $\boldsymbol{\beta}$ the vector of associated coefficients. The error term ε measures the impact of unobserved heterogeneity in patient litigiousness across hospitals. Incorporating (2) into (1) leads to overdispersion of the Poisson distribution. A mixed distribution can be obtained once an assumption is made about the distribution of exp(ε). A common assumption for the heterogeneity is the gamma distribution, and the resulting Poisson/gamma mixture can be shown to generate a negative binomial distribution for y:

$$f(y; N, \pi, \alpha) = \frac{\Gamma(\alpha^{-1} + y)}{\Gamma(\alpha^{-1})\Gamma(y+1)} \left(\frac{\alpha^{-1}}{\alpha^{-1} + N\pi} \right)^{\alpha^{-1}} \left(\frac{N\pi}{N\pi + \alpha^{-1}} \right)^{y} \tag{3}$$

where α defines the one-parameter gamma distribution for the heterogeneity variable exp(ε). The first two moments of this distribution are

$$E[y \mid N, \pi, \alpha] = N\pi \tag{4}$$

$$V[y \mid N, \pi, \alpha] = N\pi(1 + \alpha N\pi) \tag{5}$$

Using (2), and assuming data for n hospitals, we can write the log-likelihood function for (3) as

$$\ln L(\alpha, \beta) = \sum_{i=1}^{n} \{ \sum_{j=0}^{y_i - 1} \ln(j + \alpha^{-1}) - \ln y_i\,! - (y_i + \alpha^{-1}) \ln(1 + \alpha N e^{\mathbf{x_i'\beta}}) + y_i \ln \alpha + y_i(\mathbf{x_i'\beta} + \ln N) \} \tag{6}$$

Maximising (6) yields consistent estimates of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ providing that we are correct in our assumptions about the data generating process described above.

In order to test for the effect of hospital size (as measured by the annual number of FCEs, N) on the likelihood of a patient claim, it is necessary to augment the covariate vector $\mathbf{x}$. If we use the logarithm of N as a covariate, it is possible to construct a simple test by manipulation of (6). The i'th hospital's contribution to the log likelihood can be written as

$$\sum_{j=0}^{y_i - 1} \ln(j + \alpha^{-1}) - \ln y_i\,! - (y_i + \alpha^{-1}) \ln(1 + \alpha N e^{\mathbf{x_i'\beta} + \gamma \ln N}) + y_i \ln \alpha + y_i(\mathbf{x_i'\beta} + \gamma \ln N) \} \tag{7}$$

The relevant test is then against the null hypothesis that $\gamma = 1$. Rejection of this hypothesis implies that the size of the hospital has a significant impact on its litigation rate ($\gamma < 1$ implies a reduction in the litigation rate for larger hospitals, $\gamma >$

1 an increase in the litigation rate).

Results

Table 4 summarises the results from our negative binomial regressions on claim frequency. For new, outstanding, and paid claims respectively, columns 1, 3 and 5 represent the results of estimating the coefficients α, $\boldsymbol{\beta}$ and γ through maximisation of the negative binomial likelihood function specified above. In each of these equations, α was significantly different from zero, which confirms the appropriateness of the negative binomial functional form, and γ was not significantly different from 1, which implies that the size of hospitals does not appear to affect the litigation rate, after controlling for the type of hospital. Columns 2, 4 and 6 represent the same specifications augmented with the previous year's value of the dependent variable. The objective of doing this was to see if the precision of the estimates was improved by using the lagged dependent variable as a means of controlling for unobserved heterogeneity amongst hospitals. In fact the fit (measured by a pseudo R^2) was improved in each case although the qualitative results were not much affected. Overall, the results of Table 4 indicate some evidence that high excess levels reduced the observed frequency of new claims and the observed stock of outstanding claims in 2001. The evidence relating to risk management discounts is weaker, but there is a tentative suggestion that the stock of outstanding claims was lower for hospitals with high risk management discounts[9].

Table 4 here

[9] While it might be argued that both excess levels and risk management activities are endogenous, the excess levels and discounts used in our analyses were set at the beginning of the year, and can therefore be assumed to be predetermined.

Table 5 summarises our OLS results in relation to the average claim awards and costs for those hospitals with non-zero paid claims during 2001. The fit for the average award regression is very low, with no coefficients significant apart from that on the proportion of maternity admissions. This seems to indicate that differences between hospitals in the average award paid out to claimants depends mainly of the hospital's casemix, and that there is no evidence supporting the hypothesis that hospitals facing a higher proportion of costs will settle claims at lower levels. For claimant costs, there was evidence to suggest that hospitals with higher excess levels had higher costs – perhaps an indication that claims handling at the local level was not as efficient as through the NHSLA. For defence costs the only significant determinant was the previous year's average defence cost. This may imply that variations in defence costs between hospitals is a persistent but idiosyncratic factor depending on the arrangements made with local solicitors.

Table 5 here

4. Conclusions

Medical injuries can be compensated on a variety of bases: fault and causality (i.e. negligence), causality alone (i.e. no-fault or strict liability) or by virtue simply of an injury having occurred (i.e. social insurance). The main benefit of relying on a negligence rule is typically regarded as the deterrence such clearly identified liability for injury can achieve. Understanding deterrence benefits is essential feature when evaluating fault-based schemes. These benefits can be achieved through placing

responsibility on the enterprise rather than the physician, but it is important that meaningful responsibility is taken by those supplying care. Measurement of deterrence is difficult but several recent studies suggest that fault-based schemes generate more care than their no-fault counterparts. It is often hard to distinguish between appropriate and unnecessary (i.e. defensive) care but, to the extent that the latter can be reduced by clearly defined negligence standards, then the possible existence of defensive practices need not be an argument for no-fault-style reforms. The overall costs of no-fault schemes will be higher than tort if no-fault leads to less deterrence; they will be lower than under tort if no-fault reduces defensive care.

Evidence from the UK presented in this paper lends tentative support to the view that hospitals with a higher share of tort liability are more likely to take action to reduce the frequency and stock of claims. Consequently, current moves to reduce CNST excess levels to zero in the UK may have adverse consequences in terms of a higher number of claims. Efforts to replace the incentives provided by excess levels through the adoption of risk management discounts do not appear to have a strong effect. Any move to reform the current tort system by changing the basis of liability in the UK should make sure that incentives are not diluted in the search for administrative efficiency and improved access to compensation.

References

Baldwin, L-M., Hart, G., Lloyd, M. Fordyce, M. & Rosenblatt, R. A. (1995): "Defensive medicine and obstetrics", Journal of the American Medical Association, 274(20): 1606—1610.

Bovbjerg, R. R., Sloan, F. A. & Rankin, P. J. (1997): "Administrative performance of "no-fault" compensation for medical injury", Law and Contemporary Problems, 60: 71—115.

Bovbjerg, R. R. & Sloan, F. A. (1998): "No-fault for medical injury: Theory and evidence", University of Cincinatti Law Review, 67: 53—123.

Brahams, D. (1988a): "The Swedish medical insurance schemes: The way ahead for the United Kingdom?" Lancet, 1(8575-6): 43—7.

Brahams, D. (1988b): "No-fault compensation Finnish style", Lancet, 332: 733—736.

Brennan, T. A. (1987): "Untangling causation issues in law and medicine: Hazardous substance litigation", Annals of Internal Medicine, 107: 741—747.

Cummins, J. D., Weiss, M. A. & Phillips, R. D. (2001): "The incentive effects of no-fault automobile insurance", Journal of Law and Economics, 44(2): 427—464.

Danzon, P. M. (1994): "The Swedish patient compensation system: Myths and realities", International Review of Law and Economics, 14: 453—466.

Danzon, P. M. (2000a): "Liability reform: Traditional and radical alternatives", American Health Care: Government, Market Processes and the Public Interest, edited by R. D. Feldman, The Independent Institute: California.

Danzon, P. M. (2000b): "Liability for medical malpractice", Handbook of Health Economics, Volume 1B, edited by A. J. Culyer and J. P. Newhouse, Elsevier North Holland: Amsterdam.

Davis, P. B., Schug, S., Briant, R., Scott, A., Johnson, S. & Bingley, W. (2001a): "Adverse events regional feasibility study: Methodological results", New Zealand Medical Journal, 11 May: 200—202.

Davis, P. B., Lay-Yee, R., Schug, S., Briant, R., Scott, A., Johnson, S. & Bingley, W. (2001b): "Adverse events regional feasibility study: Indicative findings", New Zealand Medical Journal, 11 May: 203—205.

Davis, P. B., Lay-Yee, R., Fitzjohn, J., Hider, P., Briant, R. & Schug, S. (2001c): "Compensation for medical injury in New Zealand: Does "no-fault" increase uptake and reduce the social and clinical sensitivity of claims?", Mimeo.: University of Otago, Department of Public Health and General Practice.

Devlin, R. A. (1992): "Liability versus no-fault automobile insurance regimes: An analysis of the experience in Quebec", Contributions to Insurance Economics, edited by G. Dionne, Kluwer Academic Publishers: Norwell MA.

Farber, H. S. & White, M. J. (1994), "Medical malpractice: An empirical examination of the litigation process", RAND Journal of Economics, 22(2): 199—217.

Harvard Study (1990): Patients, Doctors and Lawyers: Medical Injury, Malpractice Litigation and Patient Compensation in New York, Report of the Harvard Medical Practice Study to the State of New York.

Horwitz, J. & Brennan, T. A. (1995): "No-fault compensation for medical injury: A case study", Health Affairs, 14: 164—179.

Hughes, J. W. & Savoca, E. (1997): "Measuring the effect of legal reforms on the longevity of medical malpractice claims", International Review of Law and Economics, 17: 261—273.

Jones, M. A. (1996): Medical Negligence, 2nd edition, Sweet and Maxwell: London.

Kessler, D. & McClennan, M. (1996): "Do doctors practice defensive medicine?", Quarterly Journal of Economics, 111(2): 353—390.

Kessler, D. & McClennan, M. (2000a): "Medical liability, managed care, and defensive medicine", NBER Working Papers Series, Working Paper 7537.

Kessler, D. & McClennan, M. (2000b): "How liability affects medical productivity", NBER Working Papers Series, Working Paper 7533.

Kochanowski, P. S. & Young, M. V. (1985): "Deterrent aspects of no-fault automobile insurance: Some empirical findings", Journal of Risk and Insurance, 52: 269—288.

Landes, E. M. (1982): "Insurance, liability and accidents: A theoretical and empirical investigation of the effects of no-fault accidents", Journal of Law and Economics, 25: 49—65.

Localio, A. R., Lawthers, A. G., Bengston, J. B., Hebert, L. E., Weaver, S. L., Brennan, T. A. & Landis, J. R. (1993): "Relationship between malpractice claims and Cesarian delivery", Journal of the American Medical Association, 269(3): 366—373.

McEwin, I. R. (1989): "No-fault and road accidents: Some Australasian evidence", International Review of Law and Economics, 9: 13—24.

Miller, R. S. (1993): "An analysis and critique of the 1992 changes to New Zealand's accident compensation scheme", Maryland Law Review, 52: 1070—1092.
Morlock, L. L. & Malitz, F. E. (1991): "Do hospital risk management programs make a difference? Relationships between hospital risk management program activities and hospital malpractice claims experience", Law and Contemporary Problems, 54: 1—22.

NZBR (1998): Accident Compensation: Options for Reform, New Zealand Business Roundtable Wellington, New Zealand; September.

Petersen, S. K. (1995): "No-fault and enterprise liability: The view from Utah", Annals of Internal Medicine, 122: 462—463.

Paterson, R. (2001): "The public's hue and cry: Medical complaints in New Zealand", Journal of Health Services Research and Policy, 6: 193—194.

Reynolds, R. A., Rizzo, J. A. & Gonzalez, M. L. (1987): "The cost of medical professional liability", Journal of the American Medical Association, 257(20): 2776—2781.

Sloan *et* al. (1997[1]): Sloan, F. A., Entman, S. S., Reilly, B. A., Glass, C. A., Hickson, G. B. & Zhang, H. H. (1997): "Tort liability and obstetricians' care levels", International Review of Law and Economics, 17: 245—260.

Sloan *et al.* (1997[2]): Sloan, F. A., Whetten-Goldstein, K., Entman, S. S., Kulas, E. D. & Stout, E. M. (1997): "The road medical injury to claims resolution: How no-fault and tort differ", Law and Contemporary Problems, 60: 35—70.

Studdert, D. M., Thomas, E. J., Zbar, B. I. W., Newhouse, J. P., Weiler, P. C., Bayuk, J. & Brennan, T. A. (1997): "Can the United States afford a 'no-fault' system of compensation for medical injury?", Law and Contemporary Problems, 60: 1—34.

Studdert, D.M. & Brennan, T. A. (2001): "No-fault compensation for medical injuries: The prospect for error prevention", Journal of the American Medical Association, 286(2): 217—223.

Weiler, P. C. (1991): Medical Malpractice on Trial, Harvard University Press: Cambridge, MA.

Weiler, P. C., Hiatt, H. H., Newhouse, J. P., Johnson, W. G., Brennan, T. A. & Leape, L. (1993): A Measure of Malpractice: Medical Injury, Malpractice Litigation and Patient Compensation, Harvard University Press: Cambridge, MA.

Whetten-Goldstein, K., Kulas, E., Sloan, F., Hickson, G. & Entman, S. (1999): "Compensation for birth-related injury: No-fault programs compared with tort system", Archives of Pediatric and Adolescent Medicine, 153, 41—48.

Zador, P. & Lund, A. (1986): "Re-analysis of the effects of no-fault auto-insurance fatal crashes", Journal of Risk and Insurance, 53: 226—241.

TABLES AND FIGURES

Table 1

Excess (£)	Freq.	Percent	Cum.
10000	159	44.54	44.54
25000	125	35.01	79.55
50000	46	12.61	92.16
100000	27	7.56	100
Total	357	100.00	

Table 2

Discount (%)	Freq.	Percent	Cum.
0	104	29.13	29.13
10	222	62.18	91.32
20	30	8.40	99.72
25	1	0.28	100.00
Total	357	100.00	

Table 3

Variable	Obs	Mean	Std. Dev.	Min	Max
FCEs	304	39451	37857	3	207764
Propn acute	357	.434	.377	0	1
Propn general	357	.137	.178	0	1
Propn maternity	357	.028	.047	0	.525

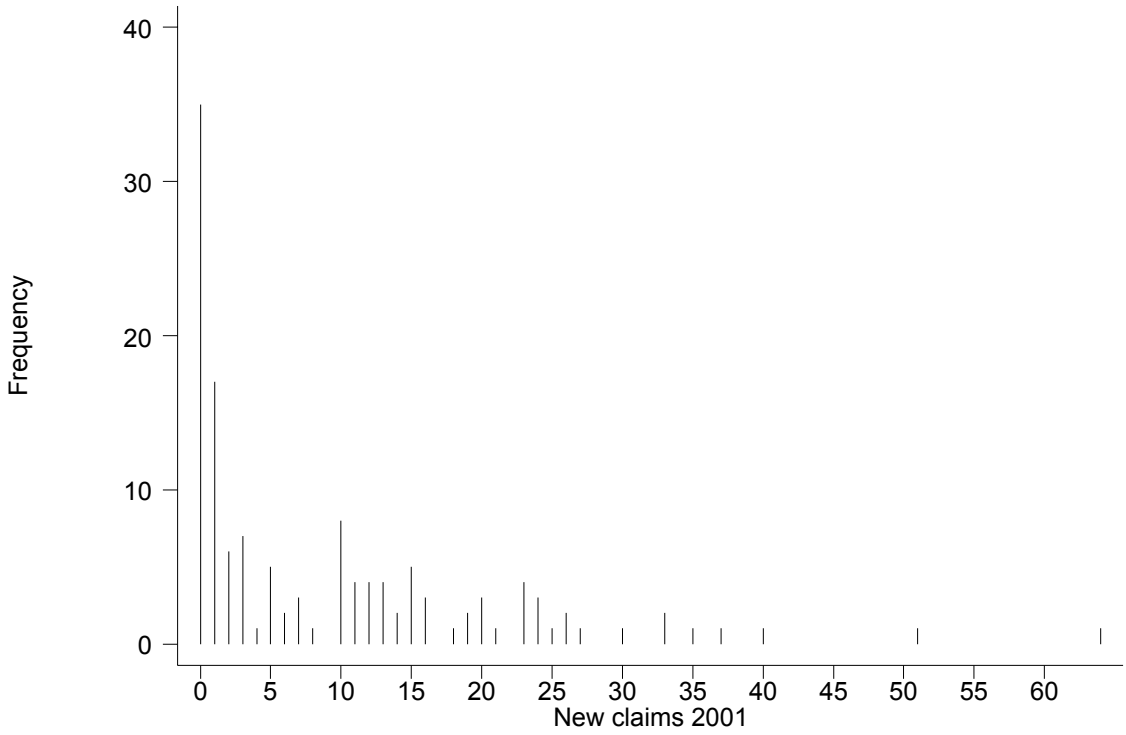
Table 4: Negative binomial regressions on clinical negligence claim frequency

	(1)	(2)	(3)	(4)	(5)	(6)
	New claims 2001		Outstanding claims 2001		Paid claims 2001	
Ln(excess) 2001	-0.120	-0.175	-0.165	-0.128	0.093	0.050
	(1.20)	(2.36)	(1.85)	(2.81)	(1.26)	(0.76)
Risk management discount 2001	-0.436	0.783	-0.490	-1.028	0.042	0.627
	(0.38)	(0.77)	(0.47)	(1.78)	(0.05)	(0.79)
Ln(FCEs) 2001	0.943	0.478	0.971	0.759	0.957	0.844
	(6.10)	(4.34)	(10.14)	(10.42)	(12.58)	(10.39)
Proportion acute	-1.132	-0.275	-0.725	0.380	-0.332	-0.327
	(1.36)	(0.90)	(0.97)	(1.58)	(1.36)	(1.32)
Proportion general	-3.221	-1.134	-2.355	-0.378	0.743	0.800
	(2.41)	(2.35)	(2.29)	(1.14)	(1.54)	(1.83)
Proportion maternity	0.150	0.648	-0.260	1.635	0.471	0.531
	(0.15)	(1.27)	(0.31)	(4.45)	(0.88)	(1.08)
New claims 2000		0.044				
		(5.11)				
Outstanding claims 2000				0.007		
				(6.66)		
Paid claims 2000						0.020
						(2.68)
Constant	-5.054	-1.713	-3.675	-3.448	-8.848	-7.535
	(4.01)	(1.55)	(3.78)	(5.11)	(11.80)	(8.54)
Observations	113	113	110	102	111	111
Pseudo R-squared	0.12	0.22	0.12	0.25	0.25	0.26
LR test of alpha=0	258	83	1421	190	65	45
Prob>Chi-squared	0.00	0.00	0.00	0.00	0.00	0.00
LR test of gamma=1	0.14		0.09		0.32	
Prob>Chi-squared	0.71		0.76		0.57	
Robust z-statistics in parentheses						

Table 5: OLS regressions on average awards and costs

	(1)	(2)	(3)
	Average award 2001	Average claimant costs 2001	Average defence costs 2001
Ln(excess) 2001	6.460	2.196	-0.608
	(1.27)	(2.37)	(0.29)
Risk management discount 2001	-48.094	22.687	27.732
	(0.48)	(1.68)	(1.54)
Proportion acute	13.603	19.710	24.989
	(0.55)	(1.43)	(1.36)
Proportion general	18.947	15.158	24.238
	(0.59)	(0.97)	(1.13)
Proportion maternity	103.143	16.458	-58.041
	(2.32)	(0.21)	(0.53)
Ln(FCEs) 2001	-6.620	-6.725	-5.786
	(1.07)	(1.92)	(1.37)
Average award 2000	-0.011		
	(0.02)		
Average claimant costs 2000		-0.219	
		(1.21)	
Average defence costs 2000			0.238
			(6.42)
Constant	13.915	39.724	54.252
	(0.19)	(1.34)	(1.33)
Observations	72	71	71
R-squared	0.05	0.18	0.17
Robust t-statistics in parentheses			

Figure 1


Frequency
40
30
20
10
0
0
5
10
15
20
25
30
35
40
45
50
55
60
New claims 2001