OPTIMAL PREVENTION WHEN INFORMAL PENALTIES MATTER:

THE CASE OF MEDICAL ERRORS

Sverre Grepperud
Institute of Health Management and Health Economics, University of Oslo and Norwegian University of Life Sciences, Aas, Norway

UNIVERSITY OF OSLO
HEALTH ECONOMICS RESEARCH PROGRAMME
Working paper 2007: 5

HERO

OPTIMAL PREVENTION WHEN INFORMAL PENALTIES MATTER:

The case of medical errors

Sverre Grepperud
Institute of Health Management and Health Economics, University of Oslo,
P.O. Box 1089 Blindern, 0317 Oslo, Norway and
Norwegian University of Life Sciences, PO 5003, N-1432, Aas, Norway.
E-mail: sverre.grepperud@medisin.uio.no

Health Economics Research Programme at the University of Oslo
HERO 2007

JEL Codes: D64, I18, K32, K42

Keywords: Iatrogenic injury, enforcement, administrative sanctions

Acknowledgements: I am indebted to V. Christiansen and K. Brekke. The usual disclaimer applies. This paper is part of the HERO program at the University of Oslo, and is funded by the Norwegian Research Council.

Health Economics Research Programme at the University of Oslo
Financial support from The Research Council of Norway is acknowledged.
ISSN 1501-9071 (print version.), ISSN 1890-1735 (online), ISBN 82-7756-177-6

Abstract

Individuals often respond with strong emotions to being penalised. Such responses suggest that informal penalties are important and play a role in creating deterrence. In this paper informal penalties are analysed in the context of medical errors. The introduction of informal penalties, if dependent upon formal ones, implies that: (i) the optimal enforcement regime becomes more lenient, and in some cases the lack of formal punishment is preferred, (ii) the first-best solution becomes unattainable, (iii) liability rates and formal penalty level are no longer perfect deterrence substitutes. In addition, powers of informal penalties provide a rationale for administrative sanctions (informal criticism, reprimands and warnings).

1. INTRODUCTION

The economic literature on deterrence and incentives has primarily focused on formal penalties with consumption implications. Formal penalties are those for which the penalizing agent is a regulatory body (governmental authority) that acts according to predetermined rules. Consumption implications refer to reductions in future consumption possibilities resulting from material or physical deprivation, e.g. fines or imprisonment. Becker's (1968) influential model on criminal offences is one example of such a perspective. Regulatory bodies (authorities), however, are not the only penalizing agents; at least two others exist – the offenders themselves and others (e.g. colleagues and peers). For such types of penalizing agents we can talk about informal penalties since they trigger negative emotions. Examples are feelings of shame, embarrassment and guilt, all a consequence of violating moral (internalized) and social norms.

The literature on how norms shape behaviour is extensive in psychology and sociology. In economics the interest has been less; however, there is now a growing literature on altruism, pro-social behaviours, reputation and intrinsic motivations, and their implications on rewards and performance incentives [see e.g. Benabou and Tirole (2003) and (2006) and the references therein]. Our analysis adds to this literature by focusing on the role informal penalties has for creating deterrence. The presence of informal penalizing agents raises the question of to what degree they dissuade individuals from certain socially unwanted acts, since they do not reduce budgetary opportunities. Questions that will be addressed are: (i) can negative feelings be utilized in a meaningful way? (ii) can such feelings replace or supplement legal sanctions and incentives? (iii) can the use of sanctions that do not affect budget constraints in a direct way, but only highlight wrongful actions, be explained by taking informal penalties into consideration? Examples of sanctions not having budgetary implications, in the following denoted administrative sanctions, are warnings, reprimands, informal criticisms, confirmations of "neglect of duty", and exclusion (loss of membership).

Some papers in economics discuss the roles self-respect and social reputation may have in creating deterrence. Brennan and Buchanan (1985) perceive legal punishment not simply as a price of an alternative course of action, but also as a confirmation of "wrongful" action, and claim that the moral dimension in itself moderates illegal behaviour. Grasmick and Bursik (1990) present a study of individuals' intention to violate the law given the perceived threats (expected penalties) for three different illegal activities (tax evasion, drunk driving and petty theft). The deterrence effect was found to be significant for legal sanctions and shame. Shame was found to be more effective than legal sanctions in the case of tax evasion, while equally effective in the case of drunk driving. Erard and Feinstein (1994) find that guilt (violating internalised values)

and shame (violating social norms) are important in explaining actual reporting behaviour in tax compliance. Gordon (1989) extends the work of Allingham and Sandmo (1972), by introducing non-pecuniary evasion costs (guilt, reduction of self-image and social stigma), to explain that some taxpayers never evade.

Administrative sanctions are frequently observed in professional associations, e.g. architects, dentists and lawyers.[1] In Nordic countries, administrative sanctions are also commonly used in the health care sector. Examples of such institutions are warnings, reprimands, informal criticism, "confirmation of misconduct" - the last being a letter from health authorities to an individual physician confirming that some degree of wrong-doing occurred in a given treatment episode.[2] Other reactions from health authorities include institutions such as "clinical supervision" and "trial-periods". Both imply some type of time-limited monitoring in response to inadequate treatment. In Denmark health care workers, primarily physicians, can be placed under the "clinical supervision" of colleagues, while Sweden uses a three-year "trial period" during which there is close follow-up by health authorities.

Negative personal experiences are common among health care personnel when it comes to medical malpractice, errors and negligence. Eldevik (2000) finds that health care workers react with surprisingly strong negative emotions to administrative sanctions. Feelings of shock and despair and emotional conflicts such as anger, depression and even suicides are reported. The emotional reactions that occur in response to material deprivation (fines, loss of authorisation, and suspension) are found to be only slightly stronger. Other studies find that significant negative reactions occur in response to process (review and litigation) and outcome (confirmation of wrong-doing). Jain and Ogden (1999) and Baker (1999) observe that general practitioners who receive a patient complaint find the experience devastating. The patient complaint appears to be a punishment in itself, regardless of the eventual decision after review. Light (1979), Marjoriebanks et al., (1996) and Hupert et al., (1996) study physicians' experiences with malpractice suits. All studies confirm the impression of strong negative feelings among health care workers in response to such experiences.

There are several reasons why feelings such as shame and guilt are frequently observed in connection with adverse events. First, the act of causing harm to others involves significant personal costs for those involved, even in the absence of any negligence, particularly if injuries happen to identifiable individuals. Second, the very intent of medical activity is to help people

[1] The disciplinary committee of the Norwegian Bar Association has the following three administrative sanctions at its disposal: (i) confirmation of misconduct, (ii) reprimands, and (iii) warnings.
[2] 1 out of 70 Norwegian practicing physicians received a "confirmation of medical misconduct" during 2005, while 1 out of 330 received a warning (Helsetilsynet, 2005).

recover from bad health, not to worsen it, a fact that may reinforce such costs. Finally, errors (injuries) may not occur intentionally, as in other harmful activities (crimes, environmental pollution and tax evasion), but accidentally. As a consequence, many physicians act in good faith.

In the following, I present an analytical model of error prevention and consider the negligence liability rule. Providers partly internalise their patients' well-being (altruism). Treatment decisions and precautionary care are collapsed into one decision (effort). Various informational imperfections are also introduced. Provider effort is unobserved by patients and imperfectly (ex-post) verified by the regulator. Patients' inability to observe health care quality explains the existence of malpractice law and other institutions dealing with quality enhancement in health care. The model follows the conventional approach in "accident models" by assuming an injury probability function decreasing in effort. Blunders, slips and misunderstandings may happen regardless of the amount of resources invested in preventing them. Physicians may err as a result of inadequate knowledge or training, and must invest in skills to reduce the probability of making mistakes (Arlen and MacLeod, 2005). In addition, court errors may happen since negligence rules are implemented under imperfect information.[3] It is generally too costly for a regulator to undertake a complete ex-ante specification of due care standards (legal standards), thus we are confronted with an incomplete contract problem.

This analysis is related to earlier work on tort liability (see e.g. Brown, 1973; Shavell, 1980 and Danzon, 1985). In addition the model draws upon the works of Becker (1968) and Polinski and Shavell (1979). Important findings in this literature are; fines are preferred over imprisonment, optimal deterrence is achieved by setting uniformly maximal penalties for all crimes, while the probability of conviction is set at the minimum necessary to enforce compliance with law. This last finding has been modified in subsequent works by including risk preferences (Polinski and Shavell, 1979), risk bearing costs among non-offenders (Kaplow, 1989), the possibility of investing in avoiding activities (Friedmann, 1981; Malik, 1990), legal expenditures by defendants and prosecutors (Rubinfeld and Sappington, 1987) and fairness and legitimacy considerations. Section 2 presents a benchmark model without informal penalties (*the conventional model*). The role of informal penalties is analysed in Section 3. Section 4 concludes and summarises the implications for policy prescription and assessment.

[3] A review under the negligence rule needs to legally establish injury, causality and substandard care relative to a standard that may be vaguely defined ex-ante. The imperfect ability to distinguish between truly careful and negligent injuries introduces the possibility of court errors.

2. A CONVENTIONAL MODEL OF ERROR PREVENTION

A medical error (adverse event), as defined in the medical literature, refers to an injury caused by medical mismanagement rather than by the underlying disease or the condition of the patient (Department of Health, 2000). The same literature often distinguishes between errors that are preventable and unpreventable relative to the current state of medical knowledge. Negligent errors represent a subset of preventable errors that satisfy certain legal criteria. The model assumes provider liability (the negligence rule), considers formal penalties (fines) and includes three possible states: the non-occurrence of medical errors, the occurrence of medical errors and the occurrence of liable medical errors. The health status of the patient, *h*, is assumed independent of preventive effort, *e*, and equal to $\hat{h}$ given the non-occurrence of an error or $\breve{h}$ for the occurrence of an error where $\hat{h} > \breve{h} > 0$. Furthermore, there is a probability $P(e)$ that the medical error occurs, where $P(0) = 1$, $\lim_{e\to\infty} P(e) \to 0$, $P_e(e) < 0$ and $P_{ee}(e) > 0$.

The occurrence of medical errors is typically not public information. Cullen et al., (1995) and Barach and Small (2000) find that the share that remains undiscovered is between 50 and 90%. Harvard Medical Practice Study (1990) finds that only 6-7% of patients suffering an injury due to negligence are believed to receive compensation. Errors that become known to the regulator do so primarily because of patient complaints. In the following the parameter $q \in [0,1]$ denotes, given the occurrence of a medical error, the share of errors for which the provider is held liable.[4] Thus *q* is the conditional probability of being detected and held liable (*liability rate*). The liability rate may depend on such factors as patients' awareness of iatrogenic injuries, the definition of due care and investments in the review and auditing processes. The penalty imposed by the regulator, $t \geq 0$, is a monetary fine (material deprivation).

Patient-physician interactions are characterised by the influence of physicians on health care use (physician agency). Formal modelling approaches of this relationship include some type of humanitarian objectives in provider utility functions rather than pure profit-maximization. Implicit treatments of agency introduce (ethical) constraints or boundaries on treatment intensities (Ma and McGuire, 1997; Iversen and Lurås, 2000), or assume that provider disutility is imposed if acting against the best interest of the patient (Dranove, 1988; McGuire and Pauly, 1991). More explicit treatments include patients' utility or health benefits as part of provider utility functions

[4] Conservative estimates from the literature find that 1% of hospital admissions result in negligent errors, see e.g. Brennan et al., (1991); Wilson et al., (1995); Thomas et al.,(1999).

(see e.g. Farley, 1986; Ellis and McGuire, 1986, 1990; Danzon, 1994 and Chalkley and Malcomson; 1998).

Here, we follow Ellis and McGuire (1990) by choosing a provider benefit function, *z*, that is the sum of the utility of net wealth, *U(A*), and patient health status. Net wealth, *A*, is the difference between *Y*, the initial level of provider wealth, and the fine, *t*, hence the maximal fine that can be imposed is $t=Y$.[5] The terms $\beta\breve{h}$ and $\beta\hat{h}$ are agency utilities, for the occurrence and non-occurrence of an error, respectively; here $\beta \in (0,1]$ denotes the degree of provider agency.[6] A strictly positive β ensures a positive level of preventive effort in the absence of any regulatory intervention. The concept of paternalistic altruism is often used to describe preferences for which specific elements of others utilities (e.g. health or health benefits) are included (Archibald and Donaldson, 1976). Thus, for $\beta < 1$ our model can be said to assume partial paternalistic altruism. The provider disutility function of preventive effort, *k(e)*, is strictly convex. The payment contract is a pure capitation contract (fixed per patient and is part of *Y)*.[7] We can now analyse the optimal behaviour of a given provider.

The provider's expected pay-off, *z*, from treating the patient is:

$$z(e) = (1-P(e))\left[U(Y) + \beta\hat{h}\right] + P(e)\left[(1-q)U(Y) + qU(Y-t) + \beta\breve{h}\right] - k(e), \qquad (1a)$$

which by simplification becomes:

$$z(e) = U(Y) - P(e)q\left[U(Y) - U(Y-t)\right] + \beta H(e) - k(e), \qquad (1b)$$

where $H(e) \equiv (1-P(e))\hat{h} + P(e)\breve{h}$. (1c)

H(e) is the expected health status of the patient, and, from former assumptions, is strictly concave in *e*; $H_e > 0$ and $H_{ee} < 0$. The following expression, for any enforcement regime {*q*,*t*}, determines optimal preventive effort, $\bar{e}$:

$$\beta H_e(\bar{e}) - P_e(\bar{e})q\left[U(Y) - U(Y-t)\right] = k_e(\bar{e}) \qquad (2)$$

Optimal preventive effort equates marginal net benefits (sum of the marginal agency utility and the marginal reduction in the expected penalty) with the marginal disutility of effort. *U(Y)-U(Y-t)*,

[5] The complete provider benefit function is as follows; $z(A, f, h) = U(A) + f + \beta h$ where *f* is the health stock of providers. Here we ignore *f* since the health of providers is assumed constant.

[6] Ellis and McGuire define β as the constant rate of substitution between wealth (net income) and benefits from treatment, where β =1 is denoted as "perfect agency". This term may be confusing since providers, contingent upon payment schedules, will trade-off patient's benefits with own net income (partial agency). Providers represent the full interest of patients only when own financial motivations are absent. Income-leisure models and pure profit maximizing providers are opposite extremes (absent agency).

[7] Pure altruistic preferences add nothing new as long co-payments are independent of provider effort.

called *the income penalty*, is the utility loss if the provider is penalised by *t.* The second order condition (s.o.c.) is available in App. A.1. Eq. (2) implicitly defines effort as function of both policy parameters: $\bar{e} = E(q,t)$, and their impacts are found by differentiating (2) with respect to *e*, *t* and *q*, which yields (see App. A.2-3);[8]

$$\frac{d\bar{e}}{dt} = E_t = -\frac{z_{et}}{z_{ee}} > 0 \tag{3a}$$

$$\frac{d\bar{e}}{dq} = E_q = -\frac{z_{eq}}{z_{ee}} > 0 \tag{3b}$$

Eqs. (3a,b) confirm the standard conclusions that preventive effort improves with higher fines and liability rates (improved deterrence). A provider invests least preventive effort (tests, precautionary care, number of visits, second opinions) when unregulated ($q = t = 0$).

To rank and compare various solutions we need to define a criteria function. The social welfare function, *S*, is defined as the unweighted sum of provider utility and patient utility minus social costs (liability costs and rehabilitation costs). The provider utility function, *z*, is defined above, while the patient benefit function, *B*, is as follows: $B(w,h) = U(w) + h$ where *w* is the initial wealth level of the patient. It follows from our specification that the patient, unlike the provider, is non-altruistic. The liability cost function is *m(q)* where $m_q > 0$ and $m_{qq} \geq 0$, and is meant to reflect audit costs, review costs and conflict resolution costs, e.g. legal costs. Although the liability rate (*q)* is a function of patient complaints, it is partly under regulator influence and can, for example, be affected by: (i) changing the rules, (ii) informational campaigns, and (iii) various other investments. Examples include changing due care standards and burdens of proof, making whistle-blowing mandatory, investing to reduce the frequency of court errors, undertaking more thorough investigations in response to patient complaints, and encouraging self-reporting by, for example, lowering the administrative costs of reporting. Rehabilitation costs, *C*, refer to costs beyond patient utility losses that follow from adverse events. Examples are investments in rehabilitation, medical expenses, new tests and longer hospital stays. Vincent et al., (2001) find that adverse events extend hospital stays by an average of eight days.[9]

Two choices in our specification of welfare function need to be clarified. The first issue concerns our inclusion of altruistic preferences, which means that health benefits now enter twice

[8] Arguments are omitted whenever doing so creates no confusion.
[9] Such costs are important but often ignored in models on medical malpractice. Kohn et al., (1999) find that in 1997 annual costs in the United States associated with preventable medical errors amounted to $17-29 billions. More than 50% of these costs were health care costs. Such costs are seldom borne by the individual providers but by hospitals and/or third-party payers (insurers).

in social welfare: first, because of patients' own evaluation (patient utilities), and second, because of providers' evaluation (agency utilities). The literature on physician agency frequently uses a different approach. Ellis and McGuire (1990) ignore patients' benefits in the physicians' objective function when defining social welfare in order to avoid "double-counting" of patient benefits. The same approach is chosen by Chalkley and Malcomson (1998) on the grounds that benevolence represents a desire to do what is in the social interest and, as such, should have no role in determining what the social interest is.[10]

Kennett (1980) refers to a particular type of altruism, genuine altruism, where the concern for others is reflected in their behavior without deriving any utility from the same behavior. According to this type of altruism, humanitarian preferences in objective functions describe behavior but should not be regarded as having effects on utility levels and thereby social welfare.[11] A somewhat different rationale for genuine altruism follows if humanitarian objectives are perceived as following from structural models of physician agency. Now altruistic preferences become reduced form formulations of physician–patient interactions i.e., portrayed as Nash bargaining games, coalition games or games of repeated interactions (see Chone´and Ma, 2004). We choose to include altruistic preferences in the welfare function as a matter of completeness; excluding altruistic preferences from the social welfare function (the genuine welfare function) is a special case of our more general formulation, and will also be discussed below.

The second issue concerns whether fines are socially neutral transfers or not. Throughout the paper I will, in order to keep the analysis as simple as possible, impose restrictions so that fines can be treated as socially neutral transfers. One simple way of doing it is to assume linear utility functions in wealth (see App. A.5-A.10 for further details).

The social welfare function, *S*, given the above assumptions, can now be expressed as:

$$S(e) = U(Y) + U(w) - p(e)C + (1+\beta)H(e) - k(e) - m(q) \qquad (4)$$

The double-counting of patients' health benefits (altruism) is captured by $(1+\beta)H(e)$. The genuine welfare function is assumed by setting $\beta = 0$ which yields lower social welfare for similar effort levels. The first–best level, $\hat{e}$ (full information solution), is derived by maximizing (4) with regard to e, which yields (the s.o.c. is presented in App. A.4):

$$(1+\beta)H_e(\hat{e}) - p_e(\hat{e})C = k_e(\hat{e}) \qquad (5)$$

The social first-order condition (see 5) deviates from the private first-order condition (see 2) in two respects. First, health improvements are valued more highly in (5). Note that this conclusion

[10] However, what is in the social interest need not be in the patient's interest.
[11] Such an assumption is somewhat controversial because it relaxes a fundamental assumption in economics - that individuals maximize their own utility.

also matters for the genuine welfare function ($\beta = 0$ in 5) as long as physicians are imperfectly altruistic ($\beta < 1$ in 2). Second, rehabilitation costs are now taken into account. Hence, the provider ignores social benefits and social costs (externalities) which become the very reason that incentives are needed to induce providers to deliver the quantity and quality of care that would have been chosen by informed patients.

In the following the problem of optimal enforcement is analyzed. The regulator is unable to observe provider effort, and a contract contingent upon health outcomes is designed. Patient complaints signal possible suboptimal care and the regulator, by the use of medical reviews, is to assess whether negligent acts were involved or not (imperfect ex-post verification). The optimal enforcement regime is derived by inserting $E(t,q)$ into S, and maximising with respect to *t* and *q*. By following this procedure, using (2), we find (see App. A.11 for the s.o.c.):

$$S_t(q,t) = E_t\left[H_e - P_e C + P_e\, q\big(U(Y) - U(Y-t)\big)\right] \tag{6a}$$

$$S_q(q,t) = E_q\left[H_e - P_e C + P_e\, q\big(U(Y) - U(Y-t)\big)\right] - m_q \tag{6b}$$

A first conclusion is that $q^* = t^* = 0$ (a non-punitive regime) cannot be part of an optimal enforcement regime since both derivatives, given former conclusions, are positive for values of *q* and *t* sufficiently close to zero. The optimal enforcement regime, given an interior solution, is derived by increasing the ratio of *q* to *t*, for $q^* > 0$ while keeping the expected income penalty, $q\big(U(Y) - U(Y-t)\big)$, constant. This procedure raises social costs, since liability costs become higher while the deterrence pressure remains unchanged, thus the optimal fine is the maximal one; $t^* = Y$ since for any $t \le Y$ it will be optimal to raise *t* and lower *q*. The optimal enforcement regime *($q^*, t^* = Y$)* satisfies the following conditions:

$$q^*\left[U(Y) - U(0)\right] < C - \frac{H_e}{P_e} \tag{7}$$

$$q^*\left[U(Y) - U(0)\right] = C - \frac{H_e}{P_e} + \frac{m_q}{P_e\, E_q} \tag{8}$$

For a maximal fine, q^* is set to balance the expected income penalty with the sum of marginal benefits and costs (externalities) and a term reflecting marginal audit costs. The optimal liability rate (q^*) increases with rehabilitation costs and marginal expected health benefits, but decreases with marginal liability costs. It is also observed that an upper corner solution cannot be ruled out since $S_q(1,Y)$ may be positive. The optimal enforcement regime described in (7) and (8) yields a

second-best solution. The reason lies with the fact that in order to induce the first-best effort level, given informational imperfections, liability costs are incurred, and under-deterrence follows $(e^* < \hat{e})$ (for proof see App. A.12-14).

Result 1: *Given social externalities and the absence of informal penalties, regulatory intervention is always optimal;* $(q^* > 0; t^* > 0)$. *In the presence of liability costs, the optimal fine is the maximal one;* $t^* = Y$.

Various special cases can be considered. Consider first the situation where the liability rate can be changed at no costs. Now, the right side of both expressions coincide and optimal enforcement is described by a menu of *t* and *q* combinations that all simultaneously fulfill (7) and (8). The two policy instruments are perfect substitutes in creating deterrence and the optimal effort level is now equal to the first-best level ($\hat{e} = e^*$; perfect deterrence). Given a genuine welfare function and keeping the assumption of no liability costs, implies that (7) can be expressed as follows;

$$q^* \left[U(Y) - U(Y - t^*) \right] = C - \frac{H_e}{P_e}(1 - \beta) \qquad (9)$$

It is observed from (9) that the optimal expected income penalty decreases with degree of agency (β). Furthermore, a punitive regime, given perfect agency ($\beta = 1$), is still in social demand because of rehabilitation costs. If, for some reason, the liability rate is fixed at a low level, it could be that (7) cannot be fulfilled and we have a third-best solution.

The above findings confirm standard conclusions about optimal enforcement in the presence of social externalities. Optimal deterrence, given costly monitoring or auditing, is achieved by maximal fines and yields a second-best solution and under-deterrence. However, our findings deviate from statements in the literature that health providers, if acting as perfect agents, should not be exposed to any malpractice pressure since the appropriate quality and effort will be provided (see e.g. Kessler and McClellan 2002a). In our model, some pressure is needed, even for a genuine welfare function, since error treatment costs are not internalised by the provider.[12]

[12] The following three conditions must be met for provider liability to be unnecessary: (i) perfect agency; (ii) fully internalised (or absent) error treatment costs, and (iii) a genuine welfare function.

3. A MODEL OF ERROR PREVENTION WITH INFORMAL PENALTIES.

In the conventional model, individuals refrain from doing something out of fear of material deprivation (the income penalty). This framework is now extended by including informal penalties, e.g. emotional and anxiety costs, imposed by the self or by colleagues and peers.[13] These penalties are viewed as utility penalties in the sense that utility, defined over initial wealth, is lowered in some states for unchanged consumption possibilities. The literature referred to in the introduction suggests that at least two groups of informal penalties are relevant for health care workers. First, reviews or litigation processes are stressful experiences for those involved, e.g. being suspected or confronted by angry patients (the process). Second, when the regulator (judicial system) confirms some type of wrongdoing by holding someone responsible, such adverse feelings can be reinforced (the outcome). The model presented below will allow for both types of informal penalties.

Furthermore, to simplify the analysis, liability costs are now set equal to zero, but the provider will still be portrayed as being altruistic (paternalistic). A natural question is whether observed negative emotions (informal penalties) among health care workers in fact follow from provider altruism and not because of reasons of self-respect and social reputation. Provider altruism is, however, too simple an explanation for several reasons: (i) strong emotional reactions also occur when patient injuries are negligible, (ii) emotional reactions differ for similar injuries, (iii) the degree to which such incidents become common knowledge plays a role, (iv) given the occurrence of patient injury, being under review or not matters, and (v) the implementation of formal penalties and their magnitude impacts the type and significance of reactions.

In the following, the informal penalties associated with each state are described. The first state, the non-occurrence of medical errors, is similar to the same state in the conventional model. For the next two states, however, informal penalties are introduced. The second state, with a conditional probability equal to *1-q*, refers to an adverse event for which the provider is not held liable. This state yields a utility equal to $V(Y) < U(Y)$, which implies that providers who experience adverse events but are not held liable, are worse off relative to not experiencing an adverse event. This assumption captures the fact that providers characterize themselves as the secondary victims, being hurt by the occurrence of medical errors. This state encompasses all types of adverse events except those for which negligence is confirmed, *i.e*., injuries caused by normal risks, undetected errors including those likely to be judged as negligent ones if they

[13] The literature sometimes apply the concept of "significant others" rather than colleagues/peers (see e.g. Grasmick and Bursik, 1990).

became public, and adverse events that trigger patient complaints but for which negligence is not confirmed. The shift in utility reflects stressful experiences due to: (i) patient complaints and litigation processes and the feelings of being suspected or being the object of others negative emotions (patients, patient family, and colleagues), (ii) various types of irrational self-reproach and ex-post regrets about ex-ante treatment strategies, and, (iii) feelings of guilt from keeping errors secret and the fear of being detected. The penalty, $U(Y)-V(Y)$, is denoted *the error penalty.*

The third state, with a conditional probability equal to q, concerns adverse events for which providers are held responsible for negligence, and yields a utility equal to $W(Y-t,t)<V(Y)\ \ \forall t$. The experienced loss of utility, $V(Y)-W(Y-t,t)$, here denoted *the negligence sanction*, contains one formal and two informal penalties. The first argument in $W(Y-t,t)$ is the conventional effect where fines reduce consumption possibilities and thus utility (the income penalty). The second argument, however, is new and reflects a separate negative effect of *t* on utility; $W_2\,(Y-t,t)<0$ (*the crowding penalty*). The crowding penalty captures the effect that the significance of informal penalties may increase with the size of the fine. Taking the derivative of $W(Y-t,t)$ with regard to the fine yields:[14]

$$\mathrm{W_t\,(Y-t,t)=-W_1\,(Y-t,t)+W_2\,(Y-t,t)<0} \qquad (10)$$

The third penalty follows since provider utility undergoes a negative shift even if no fine is imposed. This is seen when inserting for *t=0* in $W(Y-t,t)$ which yields $W(Y,0)<V(Y)$. Now, the loss of utility is $V(Y)-W(Y,0)$ (*the responsibility penalty*).[15]

The state-dependent utility function outlined above assumes four different penalties. First we have the error penalty which is an informal one. Then we have three additional ones, all being part of the negligence sanction, of which one is formal (the income penalty) and two informal. The two informal penalties (the responsibility penalty and the crowding penalty) capture the idea that holding providers responsible for errors triggers negative emotions. The significance of informal penalties is contingent upon the law itself since there is a stigma attached to being held liable. Evidence, referred to in the introduction, on health care workers' responses to administrative sanctions, e.g. reprimands and informal criticism, seems to support this assumption. Although administrative sanctions have no direct impact on consumption possibilities, they are nonetheless described as causing stressful experiences - most probably

[14] The numbered subscripts (derivatives) point to the relevant argument in *W(Y-t,t).*

[15] The model-setup does not explicitly consider court errors although the model assumptions allow for courts' inaccuracy. Type I errors, the conviction of a truly careful provider, are expected to cause significant informal penalties, but are not explicitly treated in this model.

because they confirm provider wrong-doings (blame). Being publicly charged makes providers feel insulted, disgraced and humiliated. The responsibility penalty represents negative reactions in the absence of fines. In addition, the size of the formal penalty (the fine) can be said to convey an observable signal about the degree of wrong-doing. The higher the fine, the more their competence and performance abilities have been questioned and the higher utility loss (the crowding penalty). The state-dependent utility function is still linear in wealth and the marginal utility of wealth is assumed the same across all states.

In order to describe optimal agent behavior we follow the procedure of section 2. Expected provider benefits, *r(e)*, is now obtained by replacing the utility function in the conventional model with the state-dependent utility function, thus we get;

$$r(e) = U(Y) - P(e)X + \beta H(e) - k(e) \tag{11}$$

where *H(E)* is still defined by (*1c*),

and $X = U(Y) - V(Y) + q\left[V(Y) - W(Y-t,t)\right]$ (12)

The following expression determines optimal provider effort, $\bar{e}$ (see App. B.1 for s.o.c.):

$$\beta H_e(\bar{e}) - P_e(\bar{e})X = k_e(\bar{e}) \tag{13}$$

As before (13) implicitly defines an optimal effort function:

$$\bar{e} = e(q,t)\,. \tag{14}$$

Both policy variables are positive arguments in the effort function (see App. B.2-5).

Now, comparing (13), with the same condition of the conventional model (see 2), one important difference is observed by looking at *X* (see 12). *X* is the error penalty plus *q* multiplied by the negligence sanction. Note that the parallel expression in the conventional model is *q* multiplied by the income penalty (see 2). The following expressions for *X* are derived for each of the informal penalties considered individually:

$$X^{EP} = U(Y) - V(Y) + q[V(Y) - V(Y-t)]$$

$$X^{CP} = q[U(Y,0) - U(Y-t,t)]$$

$$X^{RP} = q[U(Y) - W(Y-t)]$$

Thus X^{EP} is the expression for *X* when the only informal penalty considered is the error penalty, i.e., the crowding – and responsibility penalties are absent and so on.[16] Consequently, each of the

[16] The notation for X^i derives from the following procedures: (i) an absent crowding penalty implies that W(*Y-t,t*) is expressed as W(*Y-t*), (ii) an absent responsibility penalty implies that the W-function is replaced by the V-function, and (iii) an absent error penalty implies that the *V*-function is replaced by the *U*-function. If all three informal penalties are set equal to zero then X coincides with the expected Income penalty. This notation will be used throughout the paper when special cases are considered.

three above expressions reflects one of the informal penalties and the income penalty.

The social welfare function, *R*, still defined as the sum of physician and patient benefits minus social costs, becomes:[17]

$$R(e) = U(Y) + U(w) - P(e)C + (1+\beta)H(e) - k(e) - P(e)\left\{\left[U(Y) - V(Y)\right] + q\left[V(Y) - \Phi(Y,t)\right]\right\} \quad (15)$$

where: $\Phi(Y,t) = W(Y-t,t) + \Delta U(w,t) = W(Y-t,t) + U(w+t) - U(w)$ (16)

The last term in (15) now makes social welfare to differ in one important from social welfare of section 2 (see 4). This term is the error penalty plus the liability rate multiplied by $V(Y) - \Phi(Y,t)$. $\Phi(Y,t)$ is the social utility of wealth when a provider is fined being the sum of provider utility in the same state and the increase in patient utility that follows from the redistributed fine ($\Delta U(w,t)$; *transfer utility gain*). $V(Y) - \Phi(Y,t)$ contains two informal penalties - the responsibility penalty and the crowding penalty. This is because the formal penalty (the income penalty) and the transfer utility gain cancel each other out when fines are social neutral transfers. Furthermore, it follows from (16) that $\Phi(Y,t)$ is strictly higher than $W(Y-t,t)$ for a strictly positive *t* and equal to $W(Y-t,t)$ for *t =0* since $\Delta U(w,0) = 0$. The change in $\Phi(Y,t)$ from a higher *t*, using former assumptions, is

$$\Phi_t(Y,t) = -W_1(Y-t,t) + W_2(Y-t,t) + U_t(w+t) = W_2(Y-t,t) < 0 \quad (17)$$

It follows from (17) that $\Phi(Y,t)$ is affected by a change in *t* only if there is a crowding effect.

The first-best level, $\hat{e}$, (full information solution) is derived by maximizing (15) with respect to *e*, which yields (the s.o.c. is presented in App. B.6):

$$(1+\beta)H_e(\hat{e}) - P_e(\hat{e})C - P_e(\hat{e})\left[U(Y) - V(Y)\right] = k_e(\hat{e}) \quad (18)$$

The third term in (18), the marginal change in expected error penalty, reflects an additional social benefit from investing in preventive measures compared to the conventional model. Thus, the first-best level, defined by (18), is strictly higher than the same level defined by (5).

The optimal enforcement regime is derived by inserting (14) into (15), and maximising this expression with respect to *q* and *t* while using (13) (the s.o.c. is available in App. B.7), which yields:

$$R_t(q,t) = e_t\left[H_e - P_e C + P_e\, q\Delta U(w,t)\right] + P(e)q\Phi_t(Y,t) \quad (19a)$$

[17] As in the preceding section, a genuine welfare function implies that $\beta = 0$. The specification of the welfare function will depend on the perspective chosen. One alternative approach is to portray the social planner as the treating clinician. If so, it is debateable whether both agency preferences and the error penalty should be included.

$$R_q(q,t) = e_q\left[H_e - P_e C + P_e\, q\Delta U(w,t)\right] - P(e)\left[V(Y) - \Phi(Y,t)\right] \qquad (19b)$$

The last term in both equations (*the preference terms*), are new compared to the conventional model (see 6). Both preference terms show up because of the informal penalties (the responsibility penalty and the crowding penalty), and both are strictly negative which implies that the use of liability rates and fines incur marginal costs for society (*policy costs*).

Optimal enforcement will now be discussed in three stages. First, three possible corner solutions are considered. Second, the interior solution is analysed. Third, the interior solution is discussed for each of the three informal penalties at a time. A first conclusion is that a non-punitive regime, as an optimal policy, no longer can be ruled out as was the case in section 2. This is seen by evaluating (19b) for $q = 0$. This again implies that $t = 0$, since a positive t for $q = 0$ does not affect physician behaviour (see 13), consequently the first-order condition (19a) does not exist. Now, (19b) becomes:

$$R_q(0,0) = e_q\left[H_e - P_e C\right] - P(e)\left[V(Y) - \Phi(Y,0)\right],$$

which is strictly negative if:

$$\lim_{q\to 0} e_q\left[H_e - P_e C\right] < \lim_{q\to 0} - P(e)\left[V(Y) - \Phi(Y,0)\right]$$

The left hand side measures the marginal social benefit from a higher liability rate (q) while the right hand side is the marginal social cost from the same change. The expression in square brackets on the right hand side equals $V(Y) - W(Y,0)$ for $t=0$ which is the responsibility penalty.[18] Hence, $q^* = 0$ (no regulation) becomes increasingly likely the lower are rehabilitation costs, the lower is the marginal health benefit, and the higher is the responsibility penalty. A non-punitive regime can be optimal for an additional reason. One way to illustrate this is by ignoring the responsibility penalty - which implies that the above expression for $R_q(0,0)$ is strictly positive. In addition, (19a) now becomes:[19]

$$R_t(q,0) = e_t\left[H_e - P_e C\right] + P(e) q V_t(Y,0).$$

If $R_t(q,0) < 0\ \forall q > 0$ then $q^* = 0$. This occurs if the crowding penalty, $V_t(Y,0)$, is sufficiently negative.

We have shown that a non-punitive regime, for liability costs equal to zero, can be an optimal regulatory policy. This may happen if policy triggers significant social costs in terms of

[18] From (16) we get that $\Phi(Y,0) = W(Y,0)$ since $\Delta U(w,0) = 0$ is zero.

[19] Given an absent responsibility penalty, the *V*-function replaces the *W*-function in (17) which gives $\Phi_t(Y,0) = V_t(Y,0)$.

strong negative emotions and if such emotions induce significant behavioural responses. Furthermore, the responsibility penalty, the crowding penalty, or both can in principle make a non-punitive policy the preferred one. The error penalty, on the other hand, can not sustain the same conclusion. The reason is that the error penalty is exogenous (policy-independent). For the responsibility penalty and the crowding penalty this is not the case. Both penalties are endogenous in the sense that policy (q and t) affect their significance. We have shown that if the responsibility penalty is sufficiently high, regulatory intervention, however insignificant, may induce deterrence benefits that are outweighed by social costs (effort costs and the crowding of provider utility). Over-deterrence may also occur if the marginal crowding penalty, evaluated at *t=0*, is high. It is also observed that the second (upper) corner solution ($q^* = 1,\ t^* = Y$) can not be ruled out since both $R_t(1,Y)$ and $R_q(1,Y)$ can be strictly positive. However, such an outcome is less likely the higher the crowding penalty and the responsibility penalty, and the lower are error treatment costs and marginal health benefits.

Result 2: *Given social externalities and the presence of endogenous informal penalties, regulatory intervention need not be optimal* ($q^* = t^* = 0$). *Such a regime becomes more likely, (i) the higher the responsibility penalty,* $V(Y) - W(Y,0)$, *and, (ii) the higher the marginal crowding penalty,* $|W_2(Y-t,t)|$, *evaluated at t=0.*

An interesting observation from (13) is that providers are affected by *q* when *t=0.* This makes possible a third corner solution where $t^* = 0$ and $q^* > 0$, which reflects the case where responsibility is assigned but no material deprivation is involved. By evaluating (19a,b) for such a policy choice, the conditions for optimality become as follows:[20]

$$R_t(q,0) = e_t\left[H_e - P_e C\right] + P(e)qW_2(Y,0) < 0 \qquad (20a)$$

$$R_q(q,0) = e_q\left[H_e - P_e C\right] - P(e)\left[V(Y) - W(Y,0)\right] \geq 0 \qquad (20b)$$

The above conditions are simultaneously fulfilled if: (i) the responsibility penalty is zero or relatively insignificant (see 20b), and (ii) if imposing a fine leads to a strong negative shift in provider utility – that the marginal crowding penalty evaluated at *t=0* is significant (see 20a). The

[20] From (17) we get that $\Phi_t(q,0) = W_2(Y,0)$, while from (16) $\Phi(Y,0) = W(Y,0)$ since $\Delta U(w,0) = 0$ is zero.

role of the crowding penalty for reaching such a conclusion becomes clear when ignoring the same penalty. If so, $\Phi_t(Y,0)=0$, which implies that (20a) becomes:

$$R_t(q,0)=e_t\left[H_e-P_e C\right]<0,$$

which, according to former assumptions never can be true.

An optimal enforcement policy of the type$\{q^*>0,\ t^*=0\}$, is possible if the incurred policy costs from assigning responsibility (the responsibility penalty) are less than the social deterrence gains that follow from the same policy, on the same time as the incurred policy costs from imposing a small but positive fine are higher than social deterrence benefit. This result is interesting because it explains a frequently observed phenomenon in health care regulation – the presence of formal penalties that do not involve material deprivations (administrative sanctions). Such sanctions can now be understood as efficient institutions.[21,22]

Result 3: *Given social externalities and the presence of endogenous informal penalties, the optimal regulatory intervention may be one for which responsibility is assigned but no fine imposed* $(q^*>0;t^*=0)$. *Such an enforcement regime becomes more likely the lower the responsibility penalty,* $V(Y)-W(Y,0)$, *and the higher the marginal crowding penalty,* $\left|W_2(Y-t,t)\right|$, *evaluated at t=0.*

Now, the interior solution will be investigated. We know from the previous discussion that this solution is likely if both the responsibility penalty and the crowding penalty are modest or weak. By rewriting (19a,b) with (16) and (17), the optimality conditions (19a,b) can be presented in the following way:

$$q^*\Delta U(w,t^*)=C-\frac{H_e}{P_e}-\frac{P(e)q^*W_2(Y-t^*,t^*)}{P_e e_t} \tag{21a}$$

$$q^*\Delta U(w,t^*)=C-\frac{H_e}{P_e}+\frac{P(e)[V(Y)-W(Y-t^*,t^*)-\Delta U(w,t^*)]}{P_e e_q} \tag{21b}$$

[21] Administrative sanctions (e.g. reprimands and informal criticism) can be understood as penalties designed to create deterrence by utilizing informal penalties. An additional justification has been their role as guidance to individual providers (feedback on how to improve future clinical performance). However, the presence of negative emotions in response to such sanctions confirms that providers do not perceive them solely as pedagogical instruments. Both fines and administrative sanctions can be perceived as formal penalties (institutions) in the sense that they both are initiated by a regulatory authority, however, here the distinction between formal and informal penalties refers to the penalizing agent.

[22] An additional conclusion from the above discussion is that a higher C (rehabilitation costs), ceteris paribus, makes the two corner solutions (no regulation and administrative sanctions) less likely.

The left hand side of (21a,b) is the expected income penalty since being equal to the expected transfer utility gain. The optimal enforcement regime described in (21a,b) yields under-deterrence ($e^* < \hat{e}$; see App. B.8-11). The first-best preventive effort level ($\hat{e}$) becomes too costly to attain due to the presence of informal penalties, thus we have identified a second-best solution. The role of informal penalties is similar to the role of liability costs in the conventional model. If liability costs were included, the second-best would become a third-best.

In the following, three special cases are considered to shed more light on the mechanisms at play. In addition, such an approach will be helpful in determining the relative importance of the two policy instruments (*q* and *t*) in creating optimal deterrence. First, we consider the error penalty, ignoring the responsibility penalty and the crowding penalty. We know that $W_2\ (Y-t,t)$ in (21a) is zero when the crowding penalty is absent (see 17). Furthermore, $V(Y)-W(Y-t,t)-\Delta U(w,t)$ in (21b) is also zero when both the crowding penalty and the responsibility penalty are absent.[23] Since both preference terms in (21) are zero, there are no policy costs associated with *q* and *t*. The optimal interior solution is now characterized by:

$$q^*\Delta U(w,t^*) = C - \frac{H_e}{P_e} \tag{22}$$

The condition in (22) has similarities with the same condition for the conventional model when liability costs are ignored (se 9). However, the presence of an error penalty changes the optimal expected income penalty from $q^*[U(Y)-U(Y-t^*)]$, in the conventional model, to $q\Delta U(W,t^*) = q^*[V(Y)-V(Y-t)]$ in (22). The liability rate and the fine are now perfect deterrence substitutes since neither imposes policy costs. For the same reason, the first-best effort level is attainable. If liability costs were introduced into the model, the conclusion of section 2 would matter – a maximal penalty combined with a liability rate set at the minimum necessary to enforce optimal compliance. Note that the first-best effort level that now matters may deviate from the first-best effort in the conventional model. The presence of an error penalty, although being exogenous, introduces an additional private (and social) cost as compared to the conventional model. The risk of experiencing such a utility loss in association with adverse events will now make it more important both for providers (and society) to avoid such an outcome.

[23] An absent crowding penalty implies that *W(Y-t,t)* is replaced by *W(Y-t)*. An absent responsibility penalty implies that the *W*-function can be replaced by the *V*-function. Thus, $V(Y)-W(Y-t,t)-\Delta U(w,t)$ equals $V(Y)-V(Y-t)-\Delta U(w,t)$. By inserting for the transfer utility gain, $\Delta U(w,t)=V(Y)-V(Y-t)$, the expression becomes zero.

Now, consider the responsibility penalty only (the error penalty and the crowding penalty are absent), and the following optimality conditions:[24]

$$q^{*}\Delta U(w,t^{*}) < C - \frac{H_e}{P_e} \tag{23a}$$

$$q^{*}\Delta U(w,t^{*}) = C - \frac{H_e}{P_e} + \frac{P(e)[U(Y) - W(Y)]}{P_e\, e_q} \tag{23b}$$

From (23a) it observed that an absent crowding penalty implies that changes in *t* occur at no policy costs while from (23b) it follows that a change in *q* incurs such costs. Consequently, increasing the ratio between *q* and *t*, for $q^{*} > 0$, while keeping the expected income penalty, $q\Delta U(w,t^{*}) = q\big(W(Y) - W(Y-t)\big)$, constant, will result in higher policy costs. Hence, the optimal fine must be the maximal one, $t^{*} = Y$, while q^{*} is adjusted to make (23b) binding. This enforcement regime is similar to the one derived for the conventional model with liability costs. Consequently, liability costs and the responsibility penalty have similar implications for optimal enforcement.[25] The presence of the responsibility penalty makes *q* and *t* imperfect deterrence substitutes and *t* the preferred policy instrument.[26]

Now, consider the crowding penalty only (the error penalty and the responsibility penalty are absent), and the following optimality conditions:[27]

$$q^{*}\Delta U(w,t^{*}) = C - \frac{H_e}{P_e} - \frac{P(e)q^{*}U_2(Y - t^{*}, t^{*})}{P_e\, e_t} \tag{24a}$$

$$q^{*}\Delta U(w,t^{*}) = C - \frac{H_e^{'}}{P_e^{'}} + \frac{P(e)[U(Y^{*},0) - U(Y - t^{*}, t^{*}) - \Delta U(w,t^{*})]}{P_e^{'} e_q^{'}} \tag{24b}$$

Despite the absence of the responsibility penalty, both preference terms are negative. This means that the presence of the crowding penalty induces policy costs for both policy instruments. Their relative attractiveness now depends on their relative policy costs, e.g. a more significant

[24] An absent crowding penalty implies that $W(Y-t,t)$ is replaced by $W(Y-t)$, now $W_2(Y-t,t)$ is zero (see 21a). An absent error penalty implies that the *V*-function is replaced by the *U*-function. Now, $V(Y) - W(Y-t,t) - \Delta U(w,t)$ in (21b) equals $U(Y) - W(Y-t) - \Delta U(w,t)$. By inserting for $\Delta U(w,t) = W(Y) - W(Y-t)$ we get $U(Y) - W(Y)$ (see 23b).

[25] However, (23) differs from (7) since a significant responsibility penalty can rule out an interior solution (non-punitive regime). The optimal enforcement regime yields a second-best. Considering liability costs would induce a third-best.

[26] Another possibility is that (23a) binds while (23b) is an inequality, however, this possibility can be ruled out since it suggests an optimal policy for which *q=0* and *t>0*.

[27] An absent error penalty and an absent responsibility penalty implies that both the *V*-function and the W-function are replaced by the *U*-function. Consequently, $W_2(Y-t,t)$ in (21a) can be written as $U_2(Y-t,t)$ (see 24a) while $V(Y) - W(Y-t,t) - \Delta U(w,t)$ in (21ba) equals $U(Y,0) - U(Y-t,t) - \Delta U(w,t)$ (see 24b).

(marginal) crowding effect will change the relative attractiveness between *q* and *t*. Optimal enforcement can now, dependent on the preference terms, be characterised by various combinations of the levels of the two policy instruments.

Result 4: *Given social externalities and the presence of endogenous informal penalties, the optimal regulatory intervention may be one with a positive fine;* $(q^* > 0; t^* > 0)$. *Such an enforcement regime becomes more likely the lower the responsibility penalty,* $V(Y) - W(Y,0)$, *and the lower the marginal crowding penalty,* $|W_2(Y-t,t)|$, *evaluated at t=0. The optimal positive levels of q and t will depend on the relative significance of the responsibility penalty to the (marginal) crowding penalty.*

4. CONCLUSIONS

Strong negative sentiments in response to the imposition of penalties are common. Consequently, such emotions (informal penalties) create deterrence incentives and should be addressed when analysing optimal enforcement. In this study such an approach is undertaken. First, a conventional framework is applied to study optimal error prevention. Here, standard conclusions are confirmed: (i) some deterrence is needed since the private problem does not coincide with the social one, (ii) fines and liability rates are perfect deterrence substitutes (optimal menu) if monitoring is costless, and (iii) first-best deterrence is achieved by maximal penalties if monitoring is costly. These conclusions do not change with the two specifications of the social welfare function or the degree of provider agency.

The above findings are modified when three informal penalties are introduced. The first one, the error penalty which is independent of policy (exogenous), has minor effects on optimal deterrence. The next two, the responsibility penalty and the crowding penalty, which are endogenous in policy, change policy prescriptions in various ways and recommend less punitive enforcement regimes. Now penalties below their maximum level are found to be welfare-improving and no regulation can be an optimal policy. The first-best solution becomes unattainable, with or without liability costs, forcing us to search for second and third-best policies. Since incentive provision (delegation) comes at a cost, optimal enforcement yields under-deterrence, and fines and liability rates are imperfect deterrence substitutes. Furthermore, an optimal enforcement regime characterised by a positive liability rate and a fine equal to zero, is possible, hence administrative sanctions, institutions that assign responsibility but do not involve any material deprivations, can be understood as rational ones. A necessary condition for such a

conclusion is that informal penalties are highly sensitive to formal penalties, i.e., a small positive fine yields a significant shift in provider utility. Our conclusions may have relevance for other areas where informal penalties are important, i.e., traffic safety and workplace accidents.

The deterrent value of medical malpractice liability has been questioned because tortfeasors are typically protected by liability insurance, and because other incentives such as experience rating, mandating levels of insurance coverage and informed health care purchasers, are weak or absent. However, such presumptions ignore the role informal penalties, which are uninsurable, may have in arresting moral hazard. Furthermore, parts of the explanation for a number of recent proposals for tort reform such as strict liability, enterprise liability and no-fault systems, may lie with the significance of informal penalties (assign less blame). [28] Informal penalties may also explain a puzzle appearing in the literature: Health care workers express a significant fear of experiencing medical errors and much evidence confirms the practice of defensive medicine (see e.g. Summertone, 1995; Symon, 2000; Kessler and McClellan, 1996, 2002a,b; and Dubay et al., 1999) at the same time as formal provider liability in health care appears to be rather limited.[29]

Values of professionalism and medical ethics are promoted to protect patients and deliver high-quality care, and informal penalties, at least to some extent, become the other side of the coin. In this perspective, values established to prevent imperfect agency become the very reason that external regulation (deterrence) in pursuit of the same objectives, becomes costly. Policy-dependent informal penalties make regulation less effective, and the regulator can be said to be better off in a "conventional model" for which correct social incentives can be designed at lower costs. The significance of informal penalties can be affected by policy in other ways. They may become less important over time by campaigning against medical cultures of "naming, blaming and shaming", and more important if the use of sanctions is announced publicly, e.g. to employees, colleagues and medical associations.

An important dimension of informal penalties, as with preferences in general, i.e., risk aversion, disutility of effort and altruism, is their idiosyncratic character. Their significance varies across providers, institutions (local cultures) and, perhaps, countries (national cultures). Their impact may also vary across medical specialties due to self-selection effects among providers. This analysis recommends penalties that vary across providers according to individuals'

[28] An injured persons' right to recover damages in no-fault systems does not necessarily depend on fault alone, but also on some objective criteria.
[29] Loss of authorisation and imprisonment occur very seldom and mainly for criminal acts such as drug abuse and sexual offences, relatively rare occurrences in normal clinical practice.

sensitivity. However, the sensitivity to criticism is hidden information for regulators and will typically not be truthfully reported by providers. Hence, it becomes quite challenging to design optimal regulatory regimes, e.g. to determine when administrative sanctions become optimal institutions of deterrence. The inability to distinguish sensitive providers from insensitive ones also creates incentives among providers to exaggerate the significance of informal penalties. Such a regulatory problem could be handled by designing mechanisms inducing physicians to report their private information truthfully (revelation principle). This is an area for future research.

APPENDIX A: THE CONVENTIONAL MODEL

(I) The second order condition for the problem in (1).

$$z_{ee} = -P_{ee}q\left[U(Y) - U(Y-t)\right] + \beta H_{ee}(.) - k_{ee} < 0, \qquad \textbf{A.1}$$

It follows from former assumptions that A.1 is strictly negative.

(II) The signing of the cross partial derivatives in (3).

$$z_{et} = -P_e(e)qU_A(Y-t) > 0 \qquad \textbf{A.2}$$

$$z_{eq} = -P_e(e)\left[U(Y) - U(Y-t)\right] > 0 \qquad \textbf{A.3}$$

(III) The second order condition for the problem in (5)

$$S_{ee} = (1+\beta)H_{ee} - P_{ee}C - k_{ee} < 0 \qquad \textbf{A.4}$$

It follows from former assumptions that A.4 is strictly negative.

(IV) Revenues from fines as socially neutral transfers.

The social welfare function of section 2 is as follows;

$$S(e;q,t) = z(e;q,t) + b(e;q,t) - P(e)C - m(q) + d(e;q,t) \qquad \textbf{A.5}$$

where

$$z(e;q,t) = U(Y) - P(e)q\left(U(Y) - U(Y-t)\right) + \beta H(e) - k(e) \qquad \textbf{A.6}$$

$$b(e;q,t) = (1-P(e))\left(U(w) + \hat{h}\right) + P(e)\left(U(w) + \check{h}\right) = U(w) + H(e) \qquad \textbf{A.7}$$

$z(e;q,t)$ is the expected net benefits for the provider from a given treatment episode. $b(e;q,t)$ is the patient's expected benefits from the same episode. $d(e;qt)$ is the expected utility that follow from fines being redistributed to other members of society. Given that fines are distributed to one patient we get;

$$d(e;q,t) = P(e)q[U(w+t) - U(w)] \qquad \textbf{A.8}$$

Now, by inserting A.6-A.8 into A.5, we get:

$$S(e;q,t) = U(Y) + U(w) + (1+\beta)H(e) - k(e) - P(e)C - m(q) + P(e)qQ_i \qquad \textbf{A.9}$$

where;

$$Q = U(w+t) - U(w) + U(Y-t) - U(Y) \qquad \textbf{A.10}$$

If Q equals zero fines can be treated as socially neutral transfers. This is the case for linear utility functions (in wealth).

(V) The second order condition for the problem in (6)

The sufficient condition for a local maximum is (Sydsaeter, 1984):

$$S_{tt}S_{qq} - S_{tq}S_{qt} > 0 \qquad \textbf{A.11}$$

where;

$$S_{qq} = E_{qq}\left[H_e - P_e C + P_e q\left(U(Y) - U(Y-t)\right)\right] + \left(E_q\right)^2\left[H_{ee} - P_{ee}C + P_{ee}q\left(U(Y) - U(Y-t)\right)\right] + E_q P_e\left(U(Y) - U(Y-t)\right) - m_{qq}$$

$$S_{tt} = E_{tt}\left[H_e - P_e C + P_e q\left(U(Y) - U(Y-t)\right)\right] + \left(E_t\right)^2\left[H_{ee} - P_{ee}C + P_{ee}q\left(U(Y) - U(Y-t)\right)\right] + E_t q U_R$$

$$S_{tq} = E_{tq}\left[H_e - P_e C + P_e q\left(U(Y) - U(Y-t)\right)\right] + E_t E_q\left[H_{ee} - P_{ee}C + P_{ee}q\left(U(Y) - U(Y-t)\right)\right] + E_t P_e\left(U(Y) - U(Y-t)\right)$$

(VI) Proof of the optimal enforcement regime (eqs.7-8) inducing under-deterrence

Consider the case without agency ($\beta = 0$). Assume now that liability costs are zero (e.g. $m_q = 0$).

If, so, first-best effort ($\hat{e}$) and optimal effort (e^*) coincide. Now, the optimal enforcement regime described by eqs. (7-8) becomes:

$$q\left[U(Y) - U(Y-t)\right] = C - \frac{H_e}{P_e} \qquad \textbf{A.12}$$

Thus, a menu of $\{q,t\}$ ensures that A.12 is fulfilled. Now let $t^{**} = Y$ and set q accordingly so that A.13 is fulfilled, yielding:

$$q^{**}\left[U(Y) - U(0)\right] = C - \frac{H_e}{P_e} \qquad \textbf{A.13}$$

If now a positive marginal liability costs is assumed, $m_q > 0$, then:

$$q^{**}\left[U(Y) - U(0)\right] > C - \frac{H_e}{P_e} + \frac{m_q}{P_e E_q} \qquad \textbf{A.14}$$

In order to ensure equality in A.14**,** a value of q lower than q^{**} is necessary, which again induces an effort level (e^*) lower than the first-best level ($\hat{e}$) since from (3b) $d\bar{e}/dq > 0$.

APPENDIX B: THE MODEL WITH INFORMAL PENALTIES.

(I) The second order condition for the problem in (12)

$$r_{ee} = -P_{ee}(e)\left\{U(Y) - V(Y) + q\left[V(Y) - W(Y-t,t)\right]\right\} + \beta H_{ee} - k_{ee} < 0 \qquad \textbf{B.1}$$

It follows from former assumptions that B.1 is strictly negative.

(II) Signing the partial derivatives of e(q,t) (see 14)

The expressions for the partial derivatives of e(t,q) are:

$$e_t = -\frac{r_{et}}{r_{ee}} > 0 \qquad \textbf{B.2}$$

$$e_q = -\frac{r_{eq}}{r_{ee}} > 0 \qquad \textbf{B.3}$$

where

$$r_{eq} = -P_e\left[V(Y) - W(Y-t,t)\right] > 0 \qquad \textbf{B.4}$$

$$r_{et} = -P_e q\left[W_1 - W_2\right] > 0 \qquad \textbf{B.5}$$

(III) The second order condition for the problem in (18)

$$R_{ee} = (1+\beta)H_{ee} - P_{ee}\left[U(Y) - V(Y)\right] - P_{ee}C - k_{ee} < 0 \qquad \textbf{B.6}$$

From former assumptions it follows that B.6 is strictly negative.

(IV) The second order condition for the problem in (19)

The sufficient condition for a local maximum is (Sydsaeter, 1984):

$$R_{tt}R_{qq} - R_{tq}R_{qt} > 0 \qquad \textbf{B.7}$$

where;

$$R_{qq} = e_{qq}\left[H_e - P_e C + P_e q\left(W(Y) - W(Y-t)\right)\right] + \left(e_q\right)^2\left[H_{ee} - P_{ee}C + P_{ee}q\left(W(Y) - W(Y-t)\right)\right] + e_q P_e\left(W(Y) - W(Y-t)\right) - e_q P_e\left[V(Y) - \Phi(Y,t)\right]$$

$$R_{tt} = e_{tt}\left[H_e - P_e C + P_e q\left(W(Y) - W(Y-t)\right)\right] + \left(e_t\right)^2\left[H_{ee} - P_{ee}C + P_{ee}q\left(W(Y) - W(Y-t)\right)\right] + q\left[e_t P_e W_R + e_t P_e \Phi_t + P(e)\Phi_{tt}\right]$$

$$R_{tq} = e_{tq}\left[H_e - P_e C + P_e q\left(W(Y) - W(Y-t)\right)\right] + e_t e_q\left[H_{ee} - P_{ee}C + P_{ee}q\left(W(Y) - W(Y-t)\right)\right] + e_t P_e\left(W(Y) - W(Y-t)\right) + e_q P_e q\Phi_t + P(e)\Phi_t$$

(V) Optimal enforcement regime (see 21) and under-deterrence

We know from the discussion in section 3 that the absence of both the crowding penalty and the responsibility penalty makes optimal effort (e^*) to coincide with first-best effort ($\hat{e}$). This is because policy costs now are absent. The optimal interior solution for this case is (see 22):

$$q^*\Delta U(w,t^*) = C - \frac{H_e}{P_e} \qquad \textbf{B.8}$$

A menu of {*q*,*t*} combinations fulfil (B.8). Now, let (B.8) be fulfilled for $t^{**} = Y$ where *q* is set accordingly so that $q = q^{**}$. Thus we get:

$$q^{**}\Delta U(w,Y) = C - \frac{H_e}{P_e} \qquad \textbf{B.9}$$

In section 3 it is shown that the responsibility penalty introduces policy costs in *q*. Below we show how the introduction of the responsibility penalty may induce optimal under-deterrence. Given a responsibility penalty for $\{t^{**} = Y, q^{**}\}$, the following conditions matter:

$$q^{**}\Delta U(w,Y) = C - \frac{H_e}{P_e} \qquad \textbf{B.10}$$

$$q^{**}\Delta U(w,Y) > C - \frac{H_e}{P_e} + \frac{P(e)[V(0) - W(0,0)]}{P_e e_q} \qquad \textbf{B.11}$$

From (B.11) it follows that the responsibility penalty introduces a third term on the right hand side which is strictly negative. Hence, the right hand side becomes strictly higher than the left hand side for $\{t^{**} = Y, q^{**}\}$. In order to make (B.11) binding, *q* must be set lower than q^{**}. Since $e_q > 0$ (see B.3) optimal effort becomes lower than first-best effort ($e^* < \hat{e}$).

REFERENCES

Allingham, M.G. and A. Sandmo (1972), Income tax evasion: A theoretical analysis, Journal of Public Economics 1: 323 – 338.

Archibald, G.C. and D. Donaldson (1976), Non-paternalism and the basic theorems of welfare economics, Canadian Journal of Economics 9: 492-507

Arlen, J. and W.B. MacLeod (2005), Torts, expertise, and authority: liability of physicians and managed care organizations, RAND journal of Economics 36(3); 494-519.

Baker, R. (1999), Learning from complaints about general practitioners, British Medical Journal 318: 1567-1568.

Barach, P. and S.D. Small (2000), Reporting and preventing medical mishaps: lessons from non-medical near miss reporting systems, British Medical Journal 320: 759-763

Becker, G. (1968), Crime and punishment: an economic approach, Journal of Political Economy 76: 169-217.

Benabou, R. and J. Tirole, J (2003), Intrinsic and extrinsic motivation, Review of Economic Studies 70: 489-520.

Benabou, R. and J. Tirole, J (2006) Incentives and prosocial behavior, forthcoming American Economic Review.

Brennan, G. and J.M. Buchanan (1985), The reason for rules, Cambridge University Press, Cambridge U.K.

Brennan, T.A., L.L. Leape, N.M. Laird at al., (1991), Incidence of adverse events and negligence in hospitalised patients: results of the Harvard Medical Practice Study I, New England Journal of Medicine 324: 370-376.

Brown, J.P (1973), Toward an economic theory of liability, Journal of Legal Studies 2: 323-349.

Chalkley, M. and J.M. Malcomson (1998), Contracting for health services when patient demand does not reflect quality, Journal of Health Economics 17: 1 -19.

Chone´, P. and A. Ma (2004), Asymmetric information from physician agency: optimal payment and healthcare quantity. Draft paper. May.

Cullen D.J., D.W. Bates, J.B. Cooper, A.T. Nemeskal, and L.L. Leape (1995), The incident reporting system does not detect adverse drug events. A problem for quality improvement, Jt Comm J Quality Improvement, 21: 541-548.

Danzon, P.M. (1985), Liability and liability insurance for medical malpractice, Journal of Health Economics 4: 309-331.

Danzon, P.M. (1994), Alternative liability regimes for medical injuries: evidence from simulation analysis, The Journal of Risk and Insurance 61: 219-244.

Department of Health (2000), An organisation with a memory. Report of an expert group on learning from adverse events in the NHS chaired by the Chief Medical Officer. London: Stationary Office, 2000.

Dranove, D (1988), Demand inducement and the physician/patient relationship, Economic Inquiry 26: 281-298.

Dubay, L., Kaestner, R. and T. Waidmann, (1999), The impact of malpractice fears on cesarean section rates, Journal of Health Economics 18: 491-522.

Eldevik, M. (2000), Health care workers personal experiences - administrative sanctions. A survey: January – April 2000. The Norwegian Ministry of Health. (In Norwegian: Helsepersonellets personlige opplevelser og erfaringer ved å motta en administrativ

reaksjon. Rapport fra intervjuundersøkelse Januar – April 2000. Sosial og Helsedepartmentet.)

Ellis, R.P. and T. McGuire (1986), Provider behavior under prospective reimbursement, Journal of Health Economics 5:129-151,

Ellis R.P. and T.G. McGuire (1990), Optimal payment systems for health services, Journal of Health Economics 9: 375-396.

Erard, B. and J.S. Feinstein (1994), The role of moral sentiments and audit perceptions in tax compliance, Public Finance (supplement) 49: 70-89.

Farley, P.J. (1986), Theories of the price and quantity of physician services. A synthesis and critique, Journal of Health Economics 5: 315-333.

Friedman, D.D. (1981), Reflections on optimal punishment, or: should the rich pay higher fines?, in R.O. Zerbe (ed.), Research in Law and Economics, Greenwich, NY.JAI Press, 3: 182-205.

Gordon, J.P.F (1989), Individual morality and reputation costs as deterrents to tax evasion, European Economic Review 33: 797-805.

Grasmick, H.G. and R.B. Bursik (1990), Conscience, significant others, and rational choice: extending the deterrence model, Law and Society Review 24(3); 837-861.

Greely, H.T. (1999), Do physicians have a duty to disclose mistakes, Western Journal of Medicine 171: 82-83.

Helsetilsynet (2005), The Annual Supervision Report, the Norwegian Board of Health. www.helsetilsynet.no.

Harvard Medical Practice Study (1990), Patient, doctors and lawyers: medical injury, malpractice litigation and patient compensation in NewYork. Report of the Harvard Medical Practice Study to the State of New York.

Hupert, N., Lawthers A.G., T. Brennan, and L.M. Peterson (1996), Processing the tort deterrent signal: a qualitative study, Social Science and Medicine 43(1): 1-11.

Iversen, T. and Lurås, H. (2000), Economic motives and professional norms: the case of general medical practice, Journal of Economic Behaviour and Organization 43: 447 -470.

Jain, A. and J. Ogden (1999), General practitioners' experiences of patients' complaints: qualitative study, British Medical Journal 318: 1596 - 1599.

Kaplow, L. (1989), The optimal probability and magnitude of fines for acts that are definitely undesirable, National Bureau of Economic Research, Working Paper No. 3008.

Kennett, D.A. (1980), Altruism and economic behaviour: developments in the theory of public and private redistribution, American Journal of Economics and Sociology 39:183-198.

Kessler, D. and M. McClellan (1996), Do doctors practice defensive medicine?, Quarterly Journal of Economics 111: 353-390

Kessler, D. and M. McClellan (2002a), Malpractice law and health care reform: optimal liability policy in an era of managed care, Journal of Public Economics 84: 175-197.

Kessler, D. and M. McClellan (2002b), How liability law affects medical productivity, Journal of Health Economics 21: 931-955

Kohn, L.T. J.M. Corrigan, and M.S. Donaldson (eds.,) (1999), To err is human. Building a safer health system. Washington: National Academy Sciences.

Light, D.(1979), Uncertainty and control in professional training, Journal of Health and Social Behaviour 20:310-322.

Ma, A. and T.G. McGuire (1997), Optimal health insurance and provider payment, American Economic Review 87: 685-704.

Malik, M.S. (1990), Avoidance, screening and optimal enforcement, Rand Journal of Economics, 21(3): 341-353.

Marjoriebanks, T, M. Delveccio Good, A.G. Lawthers and L.M. Peterson, (1996), Physicians' discourse on malpractice and the meaning of medical malpractice, Journal of Health and Social Behaviour 37: 163 – 178.

McGuire T.G. and M.V. Pauly (1991), Physicians response to fee changes with multiple players, Journal of Health Economics 10: 385-410.

Polinski, M. and Shavell, S. (1979), The optimal tradeoff between the probability and magnitude of fines, American Economic Review 69: 880 – 891.

Polinski, M. and Shavell, S. (1984), The optimal use of fines and imprisonment, Journal of Public Economics 24: 89-99.

Rubinfeld, D.L. and Sappington, D.E.M. (1987), Efficient awards and standards of proof in judicial proceedings, RAND Journal of Economics 18: 308-315.

Shavell, S. (1980), Strict liability versus negligence, Journal of Legal Studies 9: 1-25.

Summertone, N. (1995), Positive and negative factors in defensive medicine: a questionnaire study of general practitioners, British Medical Journal 10; 27 – 29.

Sydsaeter, K. (1984), Matematisk Analyse II, Universitetsforlaget, Oslo. Norway.

Symon, A. (2000), Litigation and defensive clinical practice: quantifying the problem, Midwifery 16: 8 -14.

Thomas, E.J., D.M. Studdert, J.P. Newbury, B.I.W. Zbar, K.M. Howard, E.J. Williams, and T.A. Brennan (1999), Costs of medical injuries in Utah and Colorado, Inquiry 36: 255-264.

Vincent, C., G. Neale, and M. Woloshynowych (2001), Adverse events in British hospitals: preliminary retrospective record review. British Medical Journal 322(3): 517-19.

Wilson, R.M., W.B. Runciman, and R.W. Gibberd (1995), The quality in Australian healthcare study, The Medical Journal of Australia 163: 458-471.