Journal of Health Economics 9 (1990) 289–319. North-Holland

EQUITY AND ACCURACY IN MEDICAL MALPRACTICE INSURANCE PRICING*

Frank A. SLOAN and Mahmud HASSAN

Vanderbilt University, Nashville, TN 37235, USA

Received June 1989, final version received June 1990

This study examines alternative classification approaches for setting medical malpractice insurance premiums. Insurers generally form risk classification categories on factors other than the physician's own loss experience. Our analysis of such classification appproaches indicates different but no more categories than now used. An actuarially-fair premium-setting scheme based on the frequency and severity of the individual physician's losses would substantially penalize adverse experience. Alternatively, premiums could be set for groups of physicians, such as hospital medical staffs. Our simulations suggest that even staffs at rather small hospitals may be large enough to be experience-rated.

1. Introduction

Particularly in certain specialties and certain geographic areas, U.S. physicians now pay a substantial amount for medical malpractice insurance. The extremes, for example, $65,000 per year paid by neurosurgeons in Florida in 1985 for $1 million per occurrence/$1 million in the aggregate coverage are indeed newsworthy.[1] Yet medical malpractice insurers' payments on behalf of policyholders appear to be highly concentrated. Sloan et al. (1989) found, for example, that 69% of losses of anesthesiologist and obstetrician-gynecologist policyholders in Florida were incurred by 5% of such physicians during 1975–1980. Similar results were obtained for other surgical specialists in the same state. Losses for internists were even more concentrated. These results suggest that if just a few physicians could be removed from the pool, either by regulatory intervention, underwriting, or

*This research was supported in part by a grant from The Robert Wood Johnson Foundation, Grant No. 12412, 'Scheduling Damages and Experience Rating: The Next Medical Malpractice Insurance Reform.' We wish to thank Gregory Alff, Corroon and Black, Nashville, Tennessee, Kevin Thompson, Insurance Services Office, Jean Lemaire, University of Pennsylvania, and anonymous referees for advice on specific aspects of the study. We take responsibility for all errors that remain.

[1] U.S. General Accounting Office (1986, p. 33).

differential pricing, perhaps premiums to the others could be appreciably reduced.

Concentration of losses need not imply that policyholders within a risk class have different expected losses. Policyholders may be homogeneous; yet losses may be concentrated if the adverse outcome has a low probability but high associated loss. But as empirical evidence from previous studies and this study shows, the expected losses of physicians within the broad risk classes typically used by malpractice insurers vary widely.

Although, in principle, experience rating may be used to develop more accurate premiums within a risk class, in practice, use of experience rating in medical malpractice insurance has been limited. Various reasons for eschewing experience rating are given:[2]

(i) Experience rating is inappropriate for a low frequency–high severity line such as medical malpractice.
(ii) Physician policyholders would rather pay an actuarially unfair community rate.
(iii) Surcharged physicians will switch to a competing insurer willing to charge such physicians standard rates.
(iv) Fearing a surcharge, physicians will not report incidents likely to result in claims soon enough.
(v) The long lag between dates of incident and claims resolution means that experience rating would have to be based on projected losses.

Some, but not all, of these assertions are questionable.

There are several reasons why fairer and more accurate pricing would be desirable. One is equity among policyholders. A second is deterrence. Although the rationale for the tort system is to achieve an optimal level of deterrence, the combination of complete liability insurance and premium-setting independent of insureds' claims experience must blunt the deterrence incentive of the tort system.[3] Third, the very existence of insurance markets may be at stake. Unravelling may occur in competitive insurance markets [Rothschild and Stiglitz (1976)]. In view of the apparent dispersion of risks within risk classes in medical malpractice insurance, it is somewhat curious that unravelling is rare.[4]

This study examines premiums that would be set under various risk

[2]These reasons are discussed in greater detail in Sloan (1990).

[3]Medical malpractice choices only rarely contain copayment provisions [Danzon (1985, p. 129)]. Unfortunately, there is little empirical evidence on the extent to which the tort system deters accidents. Studies suggesting a deterrent effect include Boyer and Dionne (1989), Landes (1982), both for automobile liability, and Moore and Viscusi (1989) for workers' compensation.

[4]The growth of captive insurers in recent years represents a minor form of unravelling. On this phenomenon, see Bawcutt (1982).

classification schemes. In the next section, we present pertinent conceptual and empirical findings from past studies. Previous work suggests that there is considerable heteogeneity among physicians in terms of expected loss within risk categories commonly used. Thus, there may be important gains to be realized from further refinement. This study investigates alternatives with the individual physician and the hospital's medical staff as the experience-rating unit.

Section 3 describes our statistical methodology. With the individual physician as the classification unit, we consider approaches based on variables that are independent of the policyholder's actual previous claims record, that depend on the policyholder's experience, or both. An alternative classification unit is a hospital's medical staff. The attraction of using the hospital medical staff is (1) that individual physician experience counts less, and (2) by partly basing premiums of staff members on group experience, peer reviewers have a meaningful incentive to monitor their peers' work. But is this form of group rating feasible? Are many medical staffs too small for their experiences to be 'credible'? Also, to avoid bankruptcy of a hospital-based plan, would medical staffs have to put aside so much in reserves as to make this option financially unattractive? We describe the simulation methods we used to address these issues.

Section 4 presents our empirical results, first on nonexperience-based risk classification and second on experience-rated premiums – both with the individual physician and the hospital medical staff as the experience-rating unit. Our results on nonexperience-based classification suggest different, but not more, categories from those now in use. Using actuarial models described in the previous section, the premiums set on the basis of past claims frequency and severity imply substantial surcharges/discounts. The disadvantage of a steep schedule is that it may expose physicians to excessive classification risk. One way to avoid this is to combine nonexperience- and experience-rating classification approaches. The simulations designed to assess statistical credibility and initial capitalization associated with medical staffs of varying sizes suggest that setting a premium for a hospital's medical staff as a whole may be feasible for even fairly small medical staffs. The final section presents further discussion as well as the study's conclusions.

2. Background

2.1. Desirable properties of a risk classification system

The objective of risk classification in insurance is to derive categories of insureds with the same expected loss. Ideally, a risk classification system should

(1) achieve separation in expected losses across categories and achieve homogeneity within categories,
(2) be based on 'reliable' predictors (predictors not readily subject to gaming or outright falsification),
(3) provide an incentive to the policyholder for loss prevention, and
(4) not violate certain equity principles – the 'admissibility' criterion – that is, not use classification variables in violation of social norms [Shayer (1978), Abraham (1986)].

No risk classification system can be perfect on all of these criteria, and there are some inherent conflicts among them. Although some variables may be good predictors, and hence rank well in terms of separation, the cost of verifying the information may be high or unacceptably intrusive. Some variables may be good for separation, but may be 'inadmissible.' A system that tries to achieve homogeneity may no longer adequately perform the insurance function due to occasional lapses on the part of policyholders or risk classification error [Palfrey and Spatt (1985)]. To the extent that insureds are misclassified, making finer distinctions among policyholders than now would introduce new inequities. The weight to be given potential errors depends on the societal loss function. If errors are especially costly, one would want to 'flatten' premiums across risk classes. Policyholders in risk classes with larger variances in losses would generally have lower preferred rates relative to actuarial rates [Ferreira (1978)].

As stated earlier, risk classification may be based on variables that are independent of the policyholder's actual previous claims record, on the policyholder's claims experience, or a combination of the two. For analytic purposes, it is useful to consider the two types separately.

2.2. Nonexperience-based risk classification

All medical malpractice insurers distinguish among specialties and geographic areas (states or territories within a given state) for premium-setting purposes. Physicians are distinguished on the basis of broad classes of procedures performed, for example, general-family practitioners who do or do not perform surgery, and general-family practitioners and gynecologists who do or do not deliver babies.

Insurers appear to develop such risk classes from cross tabulations of historical losses [Chang and Fairley (1978), Weisberg and Tomberlin (1982)]. The use of experience from individual cells is inefficient; there is no evidence that the flexibility to be obtained from a fully interactive model achieved by such cross tabulations is worth the cost. Since observations are quickly spent with this approach, it is not surprising that insurers have been reluctant to specify more than a few risk classes. A few scholars have proposed using regression analysis for purposes of grouping policyholders into risk classes.[5]

2.3. Experience rating

Although experience rating is common in certain lines of insurance, such as workers' compensation [Victor (1985), Worrall and Butler (1988)], unemployment compensation [Topel (1984)], and automobile insurance [Lemaire (1988)], various forms of experience rating have only been applied on a limited basis in medical malpractice. A survey of 14 medical malpractice insurers conducted during 1987–1988 revealed that almost all insurers had instituted some form of experience rating at some time in the past. However, insurers charged fewer than one percent of physicians more than standard rates because of adverse claims experience at the time of the survey. Surcharges were mostly insubstantial. None exceeded 200% of standard rates, and the surcharges were often much lower than this.[6]

2.3.1. Individual physician-based experience rating

Rolph (1981) examined claims data from one company (Johnson and Higgins) to assess the desirability of using experience rating in medical malpractice. The 8,000 physicians in the sample incurred 2,300 suits. Assuming that the 2,300 suits in the data were distributed randomly, the claims frequency should be Poisson distributed with the mean (λ) equal to 2,300/8,000. Allowing λ to vary across specialty risk classes, he computed λ for each class. Using these estimated λs, Rolph computed the expected number of physicians with at least four claims over the observational period in each risk class and summed these values. The sum was 5.7 in contrast to the actual value of 46. Clearly there was heterogeneity within the risk classes. He reconfirmed these results with data from another claims file (Med-Chi).

He then specified a model that explicitly recognized heterogeneity of individual doctors within risk classes. Claims frequency was assumed to be Poisson distributed. The Poisson parameter was allowed to vary across individals within risk classes and to have a gamma distribution. Using this compound Poisson model of claims frequencies to capture heterogeneity within risk classes and estimating the parameters with method of moments, he obtained widely varying expected claims rates within risk classes. Using the same model, these results were later essentially confirmed by Nye and Hofflander (1988) with data from 1976–1984 Pennsylvania Medical Professional Liability Catastrophe Loss Fund (CAT Fund) and by Ellis et al.

[5]See Cummins et al. (1983) for an application to life insurance and Lemaire (1985) for an application to automobile insurance.

[6]See Sloan et al. (1991, ch. 8).

(1990) with published data on all claims closed in New York State [New York State Department of Public Health (1988)].

Cooil (1991) further extended research on claims frequency with unpublished data from the Florida Department of Insurance by allowing for structural changes in the underlying Poisson process over time (nonhomogeneity) and by including individual physician characteristics (such as board certification status, physician sex, country of medical school) as covariates.[7] He found that adding covariates increased the predictive power of the model, and there was some evidence of time-related structural shifts.

2.3.2. Hospital-based experience rating

Research on experience rating in medical malpractice to date has focused entirely on the individual physician as the observational unit. Alternatively, the hospital or the medical staff of the hospital could be the unit. A major potential advantage of such an approach is that it would give peer reviewers an explicit financial incentive to conduct meaningful peer review since improved quality would presumably be reflected in decreased loss and hence in reduced premiums to each group member. To the extent that physicians at a hospital were adverse to tough peer review, they would pay for their professional independence in the form of higher premiums.

Placing individual physicians in wrong groups ('risk classification error') would be less of a potential problem than under individual rating since premium-setting would be for groups rather than for individual physicians. On the other hand, cross subsidies from 'good' to 'bad apples' should be more apparent in small groups, such as a hospital staff, than in large groups, such as all physicians in a particular specialty in half a state, in which physicians have no professional relationship with one another. Groups presumably would want to exclude physicians with adverse claims experience.

Physicians forced out of groups with favorable claims records would join groups of other physicians with adverse experience. This sorting process would lead to greater homogeneity of expected loss within specialty. Further, local peer reviewers probably have an advantage in obtaining access to confidential information about the quality of individual physicians. Each group would have to allocate premium charges to individual members.

Although attractive for this reason, the suggestion also has some potential problems. Is the hospital medical staff a sufficiently large unit for experience rating purposes? This depends on the volatility of losses as a function of

[7] Data from this same source are used in the present study.

medical staff or hospital size. How could such a plan accommodate physicians with staff privileges at several hospitals? Currently, only care delivered at the hospital is subject to hospital peer review, although disciplinary actions by a licensure board or another hospital may be considered in the decision to grant admitting privileges to a physician. Would care delivered in nonhospital settings be subject to the peer review? How would the premium be financed? How would shares of individual physician members be determined? Given imperfect pricing within the group, would there be undue pressure to exclude physicians who treat patients likely to sue, for example, not because their quality of care is worse but because the population served more often produces adverse outcomes? This last question arises because the sanction to exclude physicians from the group is the ultimate sanction of hospital-based peer review. In this study, we focus on the statistical issues. Pertinent institutional issues are discussed in greater depth elsewhere.[8]

3. Methods

3.1. Data

The principal data source for this study came from the Florida Professional Liability Insurance Claims file, a computerized record of all medical malpractice claims closed in Florida. By statute, all insurers are required to report all closed medical malpractice claims to the Florida Department of Insurance. The department makes the claims information available as a public use file. Only the claimant's name is withheld; all defendants and amounts paid by insurers on their behalf are identified. We selected claims closed against physicians between January 1975 and March 1988 for this analysis. We extracted the physician's name and the indemnity and associated loss expense (largely the cost of defending the case) paid on behalf of the physician from the file. The physician's name was needed to match claims with physician characteristics from the American Medical Association (AMA) Masterfile, which contains information on all physicians. Time-invariant information is repeated on each year's Masterfile; time varying data, such as type of practice and practice hours, are collected periodically by the AMA and, when collected, are added to the most recent Masterfile. With Masterfile information, we were able to identify the universe of physicians in Florida exposed to medical malpractice claims in each year.

Closed claims data are inappropriate for purposes of assessing claims

[8]See Sloan (1990) and Sloan et al. (1991, ch. 8).

Table 1
Claims frequency and correlations between claims frequency per exposure year and claim size.

Specialty	Claims per exposure year	Correlations: claims per exposure year and claim size[a]
Family practice	0.040	0.28
General surgery	0.091	0.24
Internal medicine	0.042	0.13
Obstetrics-gynecology	0.136	0.19
Orthopedic surgery	0.186	0.15
Pediatrics	0.046	0.18
All	0.065	0.17

[a]The correlations are based on a sample that excludes physicians with no claims. All correlations are statistically significant at the one percent level. See the text for the definition of a claim used in this study.

experience. Claims closed in a given year originate from several policy years, especially in a line with a long tail such as medical malpractice insurance. We converted the data into a claim file based on the year the incident occurred.[9] The file included all claims closed with indemnity and/or associated loss expense. Claims that did not take at least a dollar to defend, thus, were excluded from the incident file and from our analysis.[10] Although all physicians were included on the file, to make the analysis manageable, we have limited this study to three specialties – internal medicine, obstetrics-gynecology, and orthopedic surgery. To be included in the study, the physician had to have practiced in Florida for at least three of the six years between 1975 and 1980.

As seen in table 1, there were appreciable differences in the total number of claims per exposure year by specialty. Claims filed per exposure year and

[9]Occurrence policies insure incidents occurring in a given year. By contrast, claims made policies insure all claims filed in a given year for incidents that occurred while the insured had coverage with the company. Although claims made coverage was prevalent during the period we studied, Florida's Department of Insurance only began collecting the date of filing in late 1985. Thus, we used the year of incident, which, for purposes of analysis of risk classification methods, is satisfactory. We constructed an incident year file for 1975–1980. Based on our own computations, it was inadvisable to extend the file beyond 1980 incident years since a number of large claims take seven, eight, or more years to close.

[10]An alternative approach would have been to limit the analysis to claims with positive indemnity or to indemnity above a threshold omitting claims with zero indemnity and positive loss expense. Even if there is no indemnity payment and hence no implied negligence, there is still a cost to policyholders of defending such claims. Some physicians may not be negligent, but may be sued frequently because of poor physician–patient relationships. Below, we distinguish between paid claims over $100,000 and those paying less amounts.

indemnity plus associated loss expense (mainly defense cost) per claim were positively correlated. There is, thus, a strong rationale for analyzing specialties separately and for analyzing claims severity as well as frequency.

3.2. Method for developing risk classification categories

The usual method for classifying risks in medical malpractice is to develop separate classes by specialty, territory (broad geographic area within a state), and sometimes whether or not a broad class of procedures are performed. We accepted the division by specialty since large differences in frequency and severity are well documented. However, it is not clear that the remaining splits are justified, and it may be possible to develop additional risk classes.

For each of the three specialties, we estimated two equations. The first was a probit equation to distinguish physicians who incurred no losses from incidents arising out of care rendered during 1975–1980 from those who incurred losses. The second was an equation for loss per exposure year for physicians with at least one dollar of indemnity and/or associated loss expense. We used a lognormal model using a maximum likelihood method developed by Amemiya (1973) to estimate the parameters of the second equation.[11] In the actuarial and statistical insurance literature on risk classification, investigators often select the set of variables offering the best prediction [Cooil (1991), Cummins et al. (1983), Lemaire (1985)]. We did not do this since, for our purposes, insignificant parameter estimates are of interest in their own right. In practical applications, it may be desirable to 'prune' the models as others have done.[12]

[11]Before selecting this model, we compared theoretical frequencies for various categories of loss with the actual values using both lognormal and gamma distributions. The fit was far better for the lognormal distribution. Often actuaries use the gamma distribution to fit loss distributions, but they use all values of loss, not just nonzero values [see Hewitt and Lefkowitz (1979) and Venter (1983)]. Not only does Amemiya's procedure take account of lognormally distributed errors, but it also corrects for heteroscedasticity, which appears to be present here. In preliminary versions, we used ordinary least squares without correcting for heteroscedasticity to estimate the second equation. The fits were far poorer than with Amemiya's method.

In this two-step specification, we treat the second outcome, the level of loss given that the physician incurred a loss as conditionally independent. In preliminary analysis, we considered a specification which allowed for the possibility that cases in the second stage were nonrandomly selected from the physician pool. On reflection, we decided that since the physicians in the second stage are conditioned on having incurred a loss by definition, a selection adjustment is inappropriate. In fact, because the equations for both stages contained the same explanatory variables, some individual coefficients were quite sensitive to inclusion/exclusion of the selection term (Mills ratio). Selection models have often been used in applied work with inadequate prior consideration about appropriate use. For recent discussions of limitations of selection adjustment approaches, see Little (1985) and Manning et al. (1987).

[12]Individual insurers often have additional information on physicians that is not in the public domain and hence would probably want to respecify the classification equations presented here.

The equations contained these explanatory variables:

(1) year of graduation from medical school – before 1938, between 1938 and 1968, and after 1968;
(2) board certification status; country of medical school – developed English-speaking, including the U.S. and Canada, developed non-English speaking, and less developed;
(3) medical school ranking, represented by a variable distinguishing graduates of the top of U.S. medical schools as ranked by the Gourman Report (1977) from graduates of other schools, both U.S. and foreign;
(4) number of direct patient care hours per week in three categories – 40 hours per week or less, 41 to 60 hours per week, and over 60 hours per week;
(5) physician gender;
(6) solo versus group practice;
(7) whether the physician engaged in research;
(8) whether the physician taught;
(9) a binary variable that took the value one if no information was available on the physician's teaching and research activity; and
(10) primary activity of the physician – medical resident, administration, research and/or teaching, or direct patient care.

The direct patient care category was further subdivided into patient care-self-employed (the omitted reference group), patient care-government, patient care-other, or patient care-unclassified. Location of practice was divided into six areas – Jacksonville standard metropolitan statistical area (SMSA), other North Florida SMSAs, Orlando and nearby SMSAs, other Central Florida SMSAs, and Miami–Ft. Lauderdale–Palm Beach SMSAs, the omitted reference group, and all nonmetropolitan (non-SMSA). When physician characteristics varied by year, we used the characteristic from the last exposure year (1975 to 1980).

For purposes of developing risk categories, we used the estimated equations to compute the expected loss for each physician in each of three specialties. We then ranked physicians in each specialty from lowest expected loss to the highest and divided the samples alternatively into four, three, and two groups. We computed the actual share of physicians with any loss and the actual loss per exposure year for each subgroup. To determine the appropriate number of groups, we required that the actual loss be statistically significant at least at the 10% level from the value in the adjacent group. In some cases, we combined groups and repeated the tests of statistical significance.

The equations were only used to assign physicians to risk categories. The statistical tests were performed on actual not predicted values. Because we used these equations to classify physicians, it should not be at all surprising that physicians in the highest group should have the highest losses. As noted earlier, a risk classification system should achieve separation in expected losses across the categories, and the purpose of this part of our analysis is to ascertain whether separation is in fact achieved.

3.3. Computation of individual experience-rated premiums

3.3.1. Theory

To derive actuarially-fair premiums based on past claims frequency, assume that the number of claims for each physician is Poisson-distributed

$$p_k(\lambda)=\frac{e^{-\lambda}\lambda^k}{k!},\quad k=0,1,\ldots, \tag{1}$$

where $p_k(\lambda)$ is the probability of observing k claims given λ, and λ is the mean claims frequency for the physician. λ follows a gamma distribution according to

$$dU(\lambda)=\frac{\tau^a e^{-\tau\lambda}\lambda^{a-1}}{\Gamma(a)}\,d\lambda, \tag{2}$$

where U is the cumulative density function of the gamma distribution, and a and τ are the two parameters of the gamma distribution (Γ), $(a,\tau)>0$. The distribution has a mean (a/τ) and a variance (a/τ^2).

The probability of observing k claims, p_k, is obtained by multiplying eqs. (1) and (2) and integrating over λ. This probability follows a negative binomial distribution,

$$p_k=\int_0^\infty \frac{e^{-\lambda(1+\tau)}\lambda^{k+a-1}\tau^a}{k!\,\Gamma(a)}\,d\lambda. \tag{3}$$

Given that the a priori distribution of λ is a gamma distribution with the parameters a and τ, then the a posteriori distribution is also a gamma distribution with the parameters $\tau'=\tau+t$ and $a'=a+k$, where t is number of periods.

It can be shown [Lemaire (1985, pp. 132–133)] that the a posteriori claim frequency of the group of policyholders with claims history $(k_1,\ldots,k_t)$ is the mean of the gamma distribution with parameters a' and τ'. The mean is

$$\lambda_{t+1}(k_1,\ldots,k_t)=(a+k)/(\tau+t) \tag{4}$$

and

$$\text{variance} = (a/\tau)(1+1/\tau). \tag{4'}$$

Up to now, we have only considered claims frequency. Especially in a high severity line such as medical malpractice, it is also desirable to incorporate the history of claim severity. Lemaire (1985) proposed a method.[13]

Lemaire distinguished 'high severity' claims from total claims. He assumed that ratio of the frequency of high severity claims (λ_c) to total claims frequency (λ) conforms to a beta distribution (β) with parameters g and h with a mean of $g/(g+h)$.

Again assume that the number of claims for each individual is Poisson distributed according to eq. (1) and λ has a gamma distribution according to (2). In addition, the ratio of severe to total claims (λ_c/λ) has a cumulative density Z. The associated probability density function is

$$\mathrm{d}Z(\lambda_c/\lambda) = \frac{(\lambda_c/\lambda)^{g-1}(1-(\lambda_c/\lambda))^{h-1}}{\lambda\beta(g,h)}\,\mathrm{d}\lambda_c. \tag{5}$$

Using the binomial theorem yields the probability that a policyholder with parameters λ, λ_c, and k claims has k_c claims (number of claims classified as severe):

$$p(k_c|k,\lambda,\lambda_c) = \frac{k!}{k_c!(k-k_c)!}\left(\frac{\lambda_c}{\lambda}\right)^{k_c}\left(1-\frac{\lambda_c}{\lambda}\right)^{k-k_c}. \tag{6}$$

By the same method used above, the a posteriori distribution for the policyholders who had the combination of k and k_c claims is the product of the gamma and beta distributions with parameters $g'=g+k_c$ and $h'=h+(k-k_c)$. The proportion of severe claims to total claims in period $t+1$ is the mean of the beta distribution, which is

$$\rho_{t+1|k,k_c} = (g+k_c)/(g+h+k). \tag{7}$$

For $t=0$,

$$\rho_1 = g/(g+h). \tag{8}$$

Under these assumptions, the proportion of high severity claims is independent of the time phasing of claim generation. As t goes to infinity, the proportion of severe claims ρ_{t+1} conditional on k and k_c tends to k_c/k.

The proportion can be expressed in terms of the standard credibility formula,

[13]Lemaire (1985, pp. 157–162).

$$\rho_{t+1|k,k_c} = (1-z)[g/(g+h)] + z(k_c/k), \tag{9}$$

with

$$z = k/(g+h+k). \tag{10}$$

At any point in time, the proportion of severe claims for an individual policyholder is a weighted average of the a priori proportion given by eq. (8) and of the one observation for the policyholder. We provide an explanation of credibility below.

3.3.2. Application

For the claims-only model, we recovered the parameters a and τ using the method of moments.[14] More specifically, we computed the mean and variance of claims frequency for each specialty assuming a gamma distribution and recovered parameters a and τ using eqs. (4) and (4′). To implement the model incorporating claims frequency and severity, we again used the method of moments to recover the parameters a and τ, which come from the gamma distribution, and g and h, which come from the beta distribution.

This treatment of high severity claims is general. The parameters g and h are the same for any definition of severe.[15] In this application, we considered paid claims with a total of indemnity plus associated loss expense in excess of \$100,000 (1987\$) to be severe. This \$100,000 threshold was at about the mean of payments for minor permanent injuries, such as loss of fingers, loss or damage to organs, and other nondisabling injuries and for wrongful death cases in Florida closed during the mid-1980s [Sloan and Hsieh (1990)]. The premium-adverse experience gradient would be steeper than the one we computed if a higher dollar threshold had been selected.[16]

We calculated experience-rated premiums as follows. Following from eq. (4), the a posteriori premium (R) equals

$$R_{t+1}(k_1, k_2, \ldots, k_t) = 100\,\frac{\tau(a+k)}{a(\tau+t)}, \tag{11}$$

where 100 is the community-based premium which a new policyholder pays (an index value). The premium is proportional to claims frequency. Recall

[14]Formulas were taken from Lemaire (1985, p. 123) and Morris and Rolph (1981, p. 306).

[15]As severity increases, g decreases. More generally, this follows from the representation of $x/(x+y)$ as a beta random variable, where x and y are independent gammas.

[16]Lemaire (1985) defined 'severe' as automobile accidents resulting in bodily injury, thus avoiding having to rely on a dollar threshold. We might have defined severe as a permanent as opposed to a temporary injury. However, the Florida Department of Insurance has only requested information on severity of injury with cases closed since October 1985.

that (a/τ) is the mean of the *prior* claims frequency. This mean was used to normalize the claims frequency actually observed for a given policyholder [see eq. (4)]. From eq. (11), if one fixes the number of claims (k) and increases the time over which they occur, the rate and hence the premium decrease. We computed premiums from (11) without adjusting for severity.

We introduced severity by calculating a proportion of severe claims (ρ) using eqs. (7) and (8). The proportion for the base period ($t=0$) was estimated for physicians in a particular class. Thus, if the class historically has had a substantial share of severe claims, this is reflected in ρ_1. To obtain an index number for severity (I_{t+1}) for updating purposes, we divided ρ_{t+1} by ρ_1 (and similarly for $\rho_{t+2}, \rho_{t+3}, \ldots$ to obtain $I_{t+2}, I_{t+3}, \ldots$). Again, ρ_1 is the initial value which is the same for all physicians in a class and ρ_{t+1} is the value for a particular policyholder based in part on his own experience at time $t+1$. We then split premiums for each k (0 claims, 1 claim, 2 claims, etc.) by multiplying I for each year by the mean premium calculated before taking account of claim severity. Thus, for example, without adjusting for claims severity, we had a premium for $k=1$; after accounting for severity, we obtained premiums for various combinations of (k, k_c) – continuing this example, for one total, none severe $(1,0)$ and one total, one severe $(1,1)$. For $k=2$, the combinations are $(2,0)$, $(2,1)$, and $(2,2)$.

3.4. Computation of hospital experience-rated premiums

3.4.1. Hospital simulations

No data are publicly available on the hospitals at which physicians have staff privileges. Therefore, we used a simulation method.

In recent years, the American Hospital Association (AHA) has collected data on the number of physicians on the hospital's active and associate medical staff by specialty as part of its Annual Survey of Hospitals. We used such data from Florida hospitals from the 1986 survey to compute mean numbers of medical staff for each of four hospital bedsize categories: under 100; 100 to 249; 250 to 399; and 400 and over. Since the number of physicians in Florida grew appreciably between 1975–1980, the years in which the malpractice incidents in the claims file occurred, and 1986, we adjusted each specialty mean downward to reflect the increase in supply between 1980 and 1986. After this adjustment, the number of physicians by hospital category were: under 100 beds – 30 physicians; 100–249 – 89 physicians; 250–399 – 181 physicians; and 400 and over – 274 physicians. The mean adjusted staff size was 150.

We then randomly selected physicians and their associated claims records (loss and associated loss expense) from our incident file. The number of physicians selected per hospital depended on the adjusted mean values of staff size and specialty mix for Florida hospitals of particular bedsizes. We drew 10 hypothetical medical staffs for each of the four bedsize categories. A given physician was drawn only once and assigned to only one of the 40 hospitals. Our matching procedure only considered numbers of physicians in various specialties. We did not adjust for possible geographic differences in staffing within Florida or other factors. Some physicians were not present in Florida for the whole six year period. When we selected a physician not present for the entire period, we drew another physician (or physicians if necessary) in the same specialty to fill in the missing years.

Once the medical staff was selected (a physician stayed on the hospital medical staff unless he left practice in Florida for one reason or another), the observed experience over the six year period for the hospital was reflected by the experience of individual physicians randomly drawn. Our method did not allow for physicians of similar risks to be attracted to the same medical staffs, which might occur if physicians of similar quality are attracted to each other; so our computations should tend to overstate differences in expected losses among members of the same staff but understate differences in losses among hospitals. Also, since each hospital in a size category was assumed to have the same number and specialty mix of physicians, this too may have led us to understate inter-hospital differences in losses. If hospital medical staff loss experience in fact differs among hospitals by more than our simulations show, the case for experience rating at the level of the hospital medical staff would be even stronger than our simulations suggest.

With the simulated losses, we addressed two issues. First, if one were to experience-rate at the level of the hospital medical staff, what would be an appropriate blend of hospital-specific loss experience and community experience? In our application, the 'community' is a hospital bedsize category in the state. If the appropriate blend involves very little hospital experience, experience rating at the level of the hospital would provide virtually no incentive to monitor physician quality. Second, if a medical staff were to form its own self-insured group, what would be the amount of capitalization required to achieve a given level of bankruptcy risk? Even a very small group could self-insure in principle, but at the price of unacceptably high initial capital contributions from individual group members.

3.4.2. Credibility

In developing classifications for premium setting purposes, actuaries face a choice between (1) having more categories and more homogeneity within categories but a less reliable estimate of the average risk for each category, and (2) fewer categories with less homogeneity within categories but a more reliable estimate of the average risk. Credibility formulae provide methods for compromising between the two extremes, that is, blending experience from individual categories with that of the categories taken together [see Morris and Van Slyke (1978)].

The appropriate blend of category-specific and pooled losses depends on the credibility factor z.[17] A weighted average premium is imputed with z_i applied to the ith category's experience and $(1-z_i)$ to the pool's experience [as in eq. (9)]. The z_i is chosen to minimize the expected sum of the squared difference between the experience-rated premium and the expected loss, where i stands for the category. The result is that for each category i,

$$z_i = \bar{L}_i/(\bar{L}_i + K), \tag{12}$$

where K is the ratio of the within-category variance of loss to the between-category variance of loss and $\bar{L}_i$ is the mean loss for the ith category for the years observed. (Both L and K are in the same units.) Intuitively, K is high when within-category losses are high relative to differences in losses among categories. With high K, the pooled experience should have a relatively high weight. In this application, the categories are medical staffs of individual hospitals and the pooled experience is for all Florida hospitals in a particular bedsize category.[18]

We computed K for each bedsize group and a K for all hospitals combined. Although traditional insurer practice in group experience rating is to compute a single K, computing Ks for each bedsize group allowed us to focus on hospitals with a particular staff size. Also, if the underlying loss distributions on an exposure-adjusted basis differs systematically by hospital size, it is desirable to distinguish among hospitals of different sizes.

3.5. Computation of initial capitalization

To assess the initial free reserves (w) needed to obtain a given level of bankruptcy risk for a self-insured group of physicians, define (suppressing subscripts), following Taylor and Buchanan (1988), a surplus process (s) as

$$s(t) = w + ct - L(t), \tag{13}$$

where c = the premium per time period and t = time period ($t = 1, \dots, n$). $L(t)$ is the loss in period t. When $s(t)$ is negative, the plan is bankrupt.

$L(t)$ has a cumulative density function $G(\cdot, t)$. The probability that $s(t) < 0$

[17] A derivation of the credibility formula is provided in Bühlmann (1967). Other sources include Hewitt (1970) and Insurance Services Office (1980).

[18] The same principle applies to eq. (10) above. The sum $(g+h)$ in eq. (10) corresponds to K in (12). The variance of the beta distribution is $gh/[(g+h)^2(g+h+1)]$ [Morris and Rolph (1981, p. 306)]. The variance is high when $(g+h)$ is low. In the individual rating case, each policyholder is his own category. Thus, the variance refers to between-category experience. So when the variance is low, the between-category differences are low and z is low. What this means in the individual rating case is that a single severe claim gets little weight if the variance is low because it is more likely to reflect a chance event rather than convey some meaningful information about the policyholder's propensity to generate severe claims.

equals the probability that the loss in year t exceeds the initial free reserve plus accumulated premium income.

$$\begin{aligned}\text{prob}\,[s(t)<0] &= \text{prob}\,[L(t)>w+ct]\\ &= 1-G(w+ct,t)=\psi(w,t),\end{aligned} \tag{14}$$

where ψ is the probability of ruin. All values of L greater than $w+ct$ imply bankruptcy. With a normal approximation to $G(\cdot)$,[19] $\Phi(\cdot)$, the probability of ruin is

$$\psi(w,t)\sim 1-\Phi\left(\frac{w+ct-\lambda t\alpha_1}{(\lambda t\alpha_2)^{1/2}}\right)=1-\Phi\left(\frac{w+\eta\lambda t\alpha_1}{(\lambda t\alpha_2)^{1/2}}\right), \tag{15}$$

where $\lambda t\alpha_1$ is the mean of $L(t)$, the product of claims frequency (λ), time (t), and average claim size (α_1), and $\lambda t\alpha_2$ is the variance of $L(t)$. The term α_2 is the variance of claim size. The term η is the markup over expected loss.[20] Then the amount of initial capital needed to ensure a probability of insolvency of no more than ε percent is given by

$$w=y_\varepsilon(\lambda t\alpha_2)^{1/2}-\eta\lambda t\alpha_1, \tag{16}$$

where y_ε is the upper 100 ε-percentile of the standard normal distribution. From eq. (16), initial capitalization is higher for a lower acceptable bankruptcy risk (lower y_ε), higher variance of loss ($\lambda t\alpha_2$), lower mean loss ($\lambda t\alpha_1$), and lower profit margin (η). Thus, a high coefficient of variation of loss implies a high w.[21] We estimated w for each of the four hospital size categories for constant y_ε (2.5%) and η (3%) and with estimated values for the other parameters of eq. (16). Since the focus of our attention centered on the volatility of loss, choice of y_ε and η made no meaningful difference to our analysis.

4. Results

4.1. Non-experience based risk classification

4.1.1. Regression results

Estimated equations for (1) the probability that the physician incurred any

[19]According to Taylor and Buchanan (1988, p. 81), similar results may be obtained using other distributions, including the gamma distribution.

[20]The markup times expected loss equals premium income (ct) minus expected loss ($\lambda t\alpha_1$).

[21]Taylor and Buchanan (1988) presented a more general version of eq. (16) which we ignored since we set premium equal to loss in the rest of our analysis.

Table 2
Probit and lognormal regressions.

Explanatory variables	Probit: Internal medicine	Probit: Orthopedic surgery	Probit: Obstetrics-gynecology	Lognormal: Internal medicine	Lognormal: Orthopedic surgery	Lognormal: Obstetrics-gynecology
Physician education characteristics						
Year of medical school graduation:						
Before 1938	−0.93[a]	−0.27	−0.28	−4,440.56[c]	16,222.9[b]	−20,037.7[a]
	(0.26)	(0.35)	(0.33)	(2,342.1)	(7,507.1)	(4,809.8)
1938 to 1968	−0.046	0.11	−0.029	−296.296	8,450.80[a]	−10,162.9[a]
	(0.12)	(0.17)	(0.14)	(494.6)	(1,069.6)	(3,231.4)
Board certified	−0.029	0.58[a]	0.57[a]	2,308.03[a]	1,921.67	−13,729.2[a]
	(0.097)	(0.19)	(0.13)	(718.0)	(1,180.4)	(3,554.9)
Medical school country:						
Developed non-English	0.17	0.59[b]	0.068	−1,184.97	−1,524.44	−2,175.13
	(0.18)	(0.29)	(0.26)	(1,199.4)	(1,995.3)	(2,732.6)
Less developed	−0.010	0.096	−0.19	2,542.09[b]	6,387.73[a]	15,175.6[a]
	(0.13)	(0.20)	(0.15)	(1,273.0)	(2,166.1)	(3,599.5)
Medical school ranking:						
Top one-third, Gourman	0.12	−0.29[b]	−0.047	3,321.37[a]	−7,464.97[a]	5,398.28[a]
	(0.10)	(0.13)	(0.12)	(728.1)	(891.50)	(918.2)
Physician practice characteristics						
Non-SMSA	−0.18	−0.070	0.11	−11,220.0[a]	−14,424.3[a]	7,341.84
	(0.20)	(0.29)	(0.21)	(1,188.0)	(2,364.3)	(6,037.7)
Jacksonville	−0.36	−0.20	−0.041	−4,580.46[a]	−1,152.42	−18,043.1[a]
	(0.22)	(0.24)	(0.18)	(1,718.7)	(4,537.1)	(1,981.2)
Orlando and nearby SMSAs	−0.059	−0.013	−0.34[b]	−5,513.57[a]	−18,264.7[a]	−15,045.7[a]
	(0.14)	(0.17)	(0.15)	(1,034.6)	(2,090.3)	(1,970.6)
Other Central Florida SMSAs	−0.072	0.18	−0.11	−6,954.70[a]	−15,268.0[a]	9,143.40[a]
	(0.11)	(0.15)	(0.14)	(1,023.3)	(2,097.3)	(3,279.0)
Other North Florida SMSAs	−0.79[a]	−0.33	−0.42[b]	−4,386.04[a]	−10,670.6[a]	−4,805.98[c]
	(0.25)	(0.23)	(0.17)	(1,293.4)	(3,077.8)	(2,528.1)
Direct patient care hours						
40 or less per week	0.10	−0.34[c]	−0.79[a]	−5,370.03[a]	−12,476.8[a]	14,586.3[a]
	(0.15)	(0.19)	(0.17)	(1,073.6)	(1,816.6)	(3,318.1)
41 to 60 per week	0.20[b]	−0.10	−0.070	−1,587.38[b]	−8,191.95[a]	4,621.87[a]
	(0.10)	(0.13)	(0.11)	(667.9)	(1,069.3)	(938.2)

Gender						
Female	−0.058	0.41	−0.13	−6,481.62[a]	–	−21,979.8[a]
	(0.23)	(1.10)	(0.24)	(1,387.6)	(–)	(3,649.3)
Solo vs. group practice	0.031	−0.044	−0.014	2,533.97[a]	−1,905.86[a]	3,211.26[a]
	(0.10)	(0.13)	(0.11)	(625.0)	(704.1)	(939.2)
Primary activity						
Government patient care	−0.43	−0.59	−0.65[b]	8,544.71[c]	−10,008.8[b]	83,648.3[b]
	(0.30)	(0.44)	(0.33)	(5,128.9)	(4,633.0)	(34,282.1)
Other patient care	−0.47[b]	−0.37	−0.66[a]	15,353.8[a]	−1,377.64	1,839.61
	(0.20)	(0.30)	(0.24)	(5,855.6)	(2,303.7)	(1,218.3)
No classification for patient care	−0.53[b]	−0.085	−0.030	2,474.58	−11,019.8[b]	8,614.25
	(0.26)	(0.33)	(0.26)	(3,607.3)	(4,972.8)	(10,700.9)
Resident	−1.44[a]	−0.55	−1.14[a]	–	−11,061.7[b]	−22,027.2
	(0.43)	(0.44)	(0.36)	(–)	(5,478.8)	(20.171.4)
Administrative	−0.64	–	−1.08[b]	–	–	–
	(0.52)	(–)	(0.55)	(–)	(–)	(–)
Research and/or teaching	−0.77[b]	−0.16	−0.86[c]	–	–	−18,848.0[a]
	(0.37)	(0.62)	(0.48)	(–)	(–)	(4,956.1)
Medical research and teaching						
No medical research	−0.088	−0.007	−0.11	5,909.93[a]	−3,166.60[b]	7,349.96[a]
	(0.16)	(0.17)	(0.18)	(852.2)	(1,613.1)	(1,212.5)
No medical teaching	0.014	−0.16	−0.024	−2,112.09[a]	−2,433.05[a]	−5,460.62[a]
	(0.097)	(0.12)	(0.11)	(791.4)	(632.7)	(918.5)
Missing research or teaching hours	−0.15	−0.90[b]	−0.58[c]	−955.47	1,918.39	43,871.5[b]
	(0.28)	(0.41)	(0.31)	(4,065.8)	(15,105.4)	(17,928.6)
Intercept	−0.72[a]	−0.19	0.059	4,422.37[a]	26,186.8[a]	30,039.1[a]
	(0.22)	(0.32)	(0.26)	(1,419.7)	(2,806.4)	(5,048.6)
Log-likelihood	−532.10	−341.67	−437.18	−918.09	−1,089.4	−1,373.6
Sample size	1,280	550	786	218	270	319

Standard errors are shown in parentheses
[a]Significant at the 1% level (two-tail test.)
[b]Significant at the 5% level (two-tail test).
cSignificant at the 10% level (two-tail test).

loss on incidents occurring during 1975–1980, and (2) the dollar value of the loss per exposure year conditional on any loss occurring are shown in table 2. Judged in terms of the number of statistically significant coefficients, the second set of regressions performs better than the first.

Since the regressions presented here are primarily used as input to the risk classification analysis that follows, we limit this discussion to principal findings.[22] First, we found no clear pattern between physicians' credentials and losses. For example, the sign on the board certification coefficient varies across regressions. Physicians with medical degrees from less developed countries (LDC) were about as likely to have any losses as other physicians, but among those with any losses, LDC doctors were likely to have larger losses per exposure year. Ranking of medical school or participation in teaching and/or research had no consistent effect on loss experience.

Second, certain other characteristics predict loss experience, but results are not consistent across specialties. For example, female obstetricians-gynecologists had lower losses than their male counterparts, holding other factors constant, but the pattern did not hold for orthopedists. (There are few females in this specialty.) Orthopedic surgeons who worked over 60 hours per week tended to have worse experience, but patterns with respect to workweek length are unclear in the other specialties. Third, in general, losses tended to be higher in South Florida, the omitted category, than in other parts of the state.[23]

4.1.2. Risk classification

Overall, results for particular explanatory variables are not sufficiently stable to allow particular variables to serve as a basis for classifying physicians for premium-setting purposes. Insurers have used region within

[22]A much more detailed analysis of factors predictive of claims experience is presented in Sloan et al. (1989). In that study, the sample was divided into four mutually-exclusive groups depending on the physician's past loss experience. The questions were estimated with multinomial logit.

[23]Cooil (1991) used the same data set to assess claims frequency in the same three specialties, but he explicitly modeled the number of claims using a modified Poisson process rather than the descriptive, aggregated probit/lognormal approach used to analyze differences in claim-no claim/loss experience in this study. Further, to study severity, Cooil selected claims with indemnities that were larger than the median of all positive claims within a specialty. By contrast, in our analysis of claims severity, we used a beta distribution with $100,000 as the minimum severe claim size (an amount appreciably above the median in all three specialties). For one specialty, we stratified the sample into thirds based on each physician's loss propensity and computed parameter estimates for each stratum using the method of moments (see table 4).

Although Cooil's final specification incorporated relatively few covariates, there are some similarities with our results. For example, in both studies, female obstetricians-gynecologists tended to have more favorable claims experience than their male counterparts. Board certified obstetrician-gynecologists had less favorable claims experience. Recent orthopedic surgeon graduates tended to have worse experience than their older colleages. Results on severity are difficult to compare. The magnitude of the parameter estimates in the Cooil study (but not their signs) differed depending on whether total claims or severe claims were analyzed.

Florida [two South Florida metropolitan counties (Dade and Broward) versus others – Nye et al. (1988)], but it is inefficient to discard information other than geographic location that is useful for purposes of risk classification.[24] The fact that coefficients differ across specialties is not a reason to discard these explanatory variables.

Some of the variables may be considered inadmissible for purposes of classification. Among these are physician gender and the physician's medical school. Such variables as physician work hours, if used to classify physicians for premium-setting purposes, may be subject to 'gaming'.[25]

Using the parameter estimates from table 2, we ranked physicians within each specialty in terms of expected loss per exposure year from lowest to highest expected loss. We then split each specialty group into as many as four equal categories and tested for differences in the mean actual loss between adjacent loss categories (table 3). To account for the skewness of the distribution of losses, we used a lognormal distribution for purposes of statistical testing.[26] When splitting the physicians into quartiles did not reveal statistically significant differences on both the fraction of cases with nonzero loss and the loss, we considered thirds and combining some of the quartiles. Although statistical significance is often lacking, the patterns are plausible, judged in terms of both the actual fraction of physicians with nonzero loss by risk category within specialty and loss per exposure year. The values for loss are low because physicians with no losses were included in the calculations.

Judging from the tests of statistical significance on actual losses, we found two risk categories for internal medicine – internists above and below the median, two for obstetrics-gynecology – physicians in the lowest third versus the others, and three categories for orthopedic surgery – physicians ranked in thirds. Thus, although the risk categories differ from those used by medical malpractice insurers in Florida, the analysis does not permit one to obtain a finer breakdown than Florida insurers use in practice. The sizable standard deviations suggest appreciable within-group heterogeneity. Hence, use of experience rating in addition to non-experience based risk classification appears to be a fruitful direction to explore.

We used the same data to

(1) estimate equations for the probability of any loss and for loss per exposure year for physicians with any loss,

[24]The Florida state legislature apparently recently recognized this point as well. The Tort Reform and Insurance Act of 1986 created a requirement that premiums reflect the number of surgical procedures performed annually by healthcare providers and claims experience.

[25]We included some variables that policymakers may find 'inadmissable'. The purpose here is to assess alternative approaches rather than produce a final 'run' for policymakers.

[26]The means and standard deviations in the table were based on untransformed data, but the statistical tests used data in log form.

Table 3
Risk classification groups.

Group		Fraction with nonzero loss	Loss ($) Mean	Standard deviation
Internal medicine				
Fourths	1	0.08	253	1,747
	2	0.13	464	2,840
	3	0.21	1,561	7,076[c]
	4	0.26	3,243	17,901
Halves*	1–2	0.11	358	2,356
	3–4	0.23[b]	2,411	13,681[b]
Thirds	1	0.08	232	1,592
	2	0.18	821	3,491
	3	0.25	3,056	16,388
All		0.17	1,386	9,873
Obstetrics-gynecology				
Fourths	1	0.27	1,538	5,781
	2	0.39	10,285	42,323[a]
	3	0.46	7,376	24,978
	4	0.54	18,416	64,780
Halves	1–2	0.33	5,934	30,557
	3–4	0.50[a]	12,925	49,434
Thirds	1	0.27	2,491	12,069
	2	0.44[a]	9,710	39,319[a]
	3	0.54[c]	15,886	57,231
	1	0.27	2,491	12,069
	2–3*	0.49[a]	12,858	49,320[a]
All		0.41	9,429	41,216
Orthopedic surgery				
Fourths	1	0.34	1,143	3,798
	2	0.47	3,139	7,418
	3	0.54	6,747	20,062
	4	0.61	16,724	44,425[a]
	1–2–3	0.45	3,684	12,749
	4	0.61[a]	16,724	44,425[a]
Halves	1–2	0.41	2,141	5,966
	3–4	0.58[a]	11,735	34,767[a]
Thirds	1	0.35	1,188	3,712
	2	0.54[a]	5,132	14,309[a]
	3	0.59	14,334	40,066[a]
All		0.49	6,956	25,421

[a]Statistically significant at 1% level (two-tail test).
[b]Statistically significant at 5% level (two-tail test).
[c]Statistically significant at 10% level (two-tail test).
*Preferred grouping for risk classification.

Table 4
Estimated parameters of gamma and beta distribution.[a]

	Internal medicine	Obstetrics-gynecology	Orthopedic surgery		
			All	Least risky one-third	Most risky one-third
a	0.29	0.44	0.55	1.21	1.56
τ	7.05	2.89	2.94	16.77	2.94
g	0.077	0.25	0.30	–	–
h	1.76	1.38	1.31	–	–

[a]The parameters are defined in section 3.2. Parameter estimates in this table were obtained using the method of moments.

(2) to classify the physicians into groups using the estimate parameters, and
(3) to test for statistical significance in differences in losses among physicians in the various groups.

In the future, it would be desirable to use new data for the second and third steps. In another study using Florida closed claims data, Sloan et al. (1989) found that physicians with adverse claims experience from incidents that arose between 1975 and 1980 had appreciably worse claims experience from incidents that arose during 1981 to 1983. Thus, there is evidence for stability in relative loss experience over time.

4.2. *Experience rating with the individual physician as the experience-rating unit*

We computed values for a, τ, g, and h for each specialty using the method of moments (table 4). The parameters are similar for the two surgical specialties. Comparatively, the variability of the gamma distribution of claims of the internists is less dispersed; that is, there appears to be less heterogeneity of claims experience within this specialty [see eq. (4′) above].

Using eq. (11), we then computed experience-rated premiums (table 5). Such premiums can be established without (part I) or with risk classification (part II). We only show cells containing at least one percent of physicians (in a row).

Past experience has a substantial effect on premiums charged. Physicians with good records receive large discounts. Conversely, an increased number of claims leads to surcharges as large as 600%. The gradient is comparatively high for internal medicine because having a large number of claims is unusual in this field. The discount-surcharge schedule shown is particularly severe when one considers that all claims resulting in some payment of indemnity-defense cost are counted. Thus, for example, if an insurer paid $100 in defending an internist's claim in year 1, the internist's premiums would quadruple. To be premium neutral, premium decreases from discounts

Table 5
Experience-rated premiums by year.

Year	Number of claims 0	1	2	3	4	5	6	7
	Part I. Premiums without risk classification[a]							
Internal medicine								
0	100(100)							
1	88 (97)	392 (2)	696 (1)					
2	78 (94)	348 (5)	619 (1)					
3	70 (91)	314 (7)	557 (2)					
4	64 (86)	285(11)	507 (2)					
5	59 (83)	262(13)	465 (3)	668(1)				
6	54 (80)	242(15)	429 (4)	617(1)				
Obstetrics-gynecology								
0	100(100)							
1	74 (91)	241 (8)	408 (1)					
2	59 (81)	192(15)	325 (3)	458(1)				
3	49 (72)	159(19)	270 (7)	380(2)				
4	42 (64)	136(22)	230 (8)	325(3)	419(1)	513(1)	607(1)	
5	37 (56)	119(24)	201(11)	284(6)	366(1)	448(1)	530(1)	613(1)
6	32 (48)	106(27)	179(13)	252(6)	325(2)	398(2)	471(1)	544(1)
Orthopedic surgery								
0	100(100)							
1	75 (88)	210(10)	345 (2)					
2	60 (76)	167(15)	275 (6)	383(3)				
3	50 (67)	139(19)	229 (9)	318(3)	408(2)			
4	42 (60)	119(22)	196(13)	273(3)	349(2)			
5	37 (53)	104(25)	171(12)	238(5)	305(2)	372(2)	439(1)	
6	33 (50)	92(24)	152(15)	212(6)	271(1)	331(2)	390(1)	450(1)
	Part II. Premiums with risk classification							
Orthopedic surgery – least risky one-third								
0	100							
1	94	172	250					
2	89	163	237	311				
3	85	155	225	295	365			
4	81	148	214	281	348			
5	77	141	204	268	332	396		
6	74	135	195	256	317	378		
Orthopedic surgery – most risky one-third								
0	100							
1	75	122	170					
2	60	98	136	174				
3	50	81	113	145	176			
4	42	70	97	124	151			
5	37	61	85	108	132	156		
6	33	54	75	96	117	138		

[a]Numbers in parentheses are percentage of physicians in the cell by row.

should offset increases from surcharges [Lemaire (1985)]. The premiums in table 5 do not quite show such balance (the schedule generates too much money over time) because of the nonstationarity of claims frequency during the observational period [Cooil (1991)]. In actual practice, it would be necessary to correct for this factor.

Combining experience rating and risk classification substantially reduces the effect of prior experience on premiums. Orthopedic surgeons classified in the bottom third would pay 12 times the premium of orthopedists in the top third initially (table 3). However, since loss is anticipated for this group, even having five claims in six years only results in a surcharge of 38%. However, if an orthopedist in the top third had such claims experience, then he would face a surcharge of nearly 300%. In the sixth year, he would pay about the same premium as a claimless orthopedist who was initially classified in the bottom third.

We also used eqs. (7) and (11) to compute experience-rated premiums based on claims severity and frequency (see table 6). Not surprisingly, having a large claim results in a huge surcharge.

As noted above, to the extent that risk classification error is seen as costly, it would be desirable to flatten premiums across risk classes. Use of nonexperience-based risk classification in combination with experience rating is a step in this direction. Whether the schedule based on this approach is flat enough depends on the societal loss function. The cost of misclassifying is unlikely to be so high as to justify ignoring a physician's prior loss experience.

4.3. Experience rating with the hospital medical staff as the experience-rating unit

To determine the optimal blend of hospital-specific and community-based experience, we computed the credibility factor (z_i) from eq. (12) (table 7). Estimated values of K (ratio of within-category of loss to between-category variance of loss) vary from a low of \$2.4 thousand (100 to 249 bedsize) to a high of \$11.3 thousand (400 and over beds). Overall, K is less than one thousand dollars. The low values of K reflect substantial differences in losses among hospitals observed over a six-year period, even within bedsize categories. If we had been able to base the calculations on actual physician affiliations with hospitals, the differences in mean losses among hospitals would have probably been greater.

Given these low values of K, almost all weight is given to the individual hospital's experience (table 7). For example, for under 100 bed hospitals, the mean z_i is 0.88. The hospital with the worst experience (Hospital No. 1) would pay an annual premium of over \$152,000 ($0.88 \times \$165.73 + 0.12 \times \$55.28$), which amounts to a surcharge of almost 200% over the community rate; the

Table 6
Severity-adjusted experience-rated premiums by year.[a]

k	0	1	1	2	2	2	3	3	3	4	4	4	4	5
k_c	0	0	1	0	1	2	0	1	2	0	1	2	3	0
Year						Internal medicine								
0	100													
	(100)													
1	88	254	3,570	333										
	(97)	(2)	(*)	(*)										
2	78	226	3,175	296										
	(94)	(5)	(*)	(1)										
3	70	203	2,860	267										
	(91)	(7)	(*)	(2)										
4	64	185	2,601	243										
	(86)	(10)	(1)	(2)										
5	59	169	2,385	223	3,133	6,044	254							
	(83)	(12)	(1)	(3)	(*)	(0)	(1)							
6	54	157	2,202	206	2,893	5,581	234							
	(80)	(14)	(1)	(3)	(1)	(0)	(1)							
Year						Obstetrics-gynecology								
0	100													
	(100)													
1	74	149	749	183										
	(91)	(7)	(1)	(1)										
2	59	119	594	146										
	(81)	(14)	(1)	(3)										
3	49	99	492	121	604	1,087	134	667						
	(72)	(18)	(1)	(6)	(1)	(0)	(2)	(*)						
4	42	84	421	103	516	929	114	570	1,027	121				
	(64)	(21)	(2)	(7)	(2)	(0)	(3)	(*)	(0)	(1)				
5	37	74	368	90	451	812	100	498	896	106				
	(56)	(22)	(2)	(10)	(2)	(0)	(4)	(1)	(*)	(1)				
6	32	65	326	80	400	720	89	442	796	94	469			
	(48)	(25)	(2)	(11)	(2)	(0)	(4)	(1)	(*)	(1)	(1)			
Year						Orthopedic surgery								
0	100													
	(100)													
1	75	129	558	153										
	(88)	(9)	(1)	(2)										
2	60	103	445	122	529									
	(76)	(14)	(1)	(4)	(2)									
3	50	86	370	102	440	779	111	480	848	117	505	893	1,281	
	(67)	(17)	(2)	(8)	(1)	(*)	(2)	(*)	(0)	(1)	(1)	(*)	(1)	
4	42	73	317	87	377	666	95	411	726	100	432	765	1,097	
	(60)	(20)	(2)	(11)	(2)	(0)	(2)	(1)	(0)	(1)	(*)	(*)	(*)	
5	37	64	277	76	329	582	83	359	635	87	378	668	959	91
	(53)	(23)	(2)	(10)	(2)	(0)	(4)	(1)	(*)	(1)	(0)	(*)	(0)	(1)
6	33	57	246	68	293	517	74	319	564	78	336	594	851	80
	(50)	(23)	(1)	(13)	(2)	(0)	(4)	(1)	(*)	(0)	(*)	(*)	(0)	(2)

[a]Numbers in parentheses are the percent of physicians in row with the combination of k and k_c. The asterisk signifies that there were some physicians in the category but fewer than one percent of the physicians fell into the category.

Table 7

Results of hospital simulations (thousands of 1987\$, variance in 1987\$).

	Beds < 100		100 ≦ Beds < 250		250 ≦ Beds < 400		Beds ≧ 400	
Hospital	Loss (L)	Variance of L	Loss (L)	Variance of L	Loss (L)	Variance of L	Loss (L)	Variance of L
1	165.7	116,458	398.1	117,012	618.3	136,441	618.2	157,595
2	15.9	299	129.6	11,762	730.8	198,373	393.3	153,732
3	9.3	421	389.6	133,271	736.3	941,758	596.4	259,563
4	4.6	59	44.6	1,180	792.9	958,232	889.0	345,605
5	39.2	5,752	107.4	5,804	712.3	210,606	744.1	367,990
6	45.2	3,251	63.9	1,643	146.5	7,286	461.9	71,946
7	80.9	25,978	74.6	5,130	476.6	308,012	796.7	339,925
8	29.5	1,532	185.3	38,640	337.4	57,610	700.9	175,042
9	68.6	4,913	176.7	22,781	174.8	37.607	599.5	396,973
10	93.9	20,176	122.3	44,269	343.8	145,737	519.3	381,660
Mean per hospital	55.3	17,884	169.2	47,016	506.9	308,850	631.9	246,098
Mean per physician	1.8		1.9		2.8		2.3	
K	7.39		2.38		5.04		11.27	
$\bar{z}_i$	0.882		0.986		0.990		0.982	
w	260.45		377.75		1,058.62		990.01	

hospital with the most favorable experience (Hospital No. 4) would pay almost $11,000. This pattern of results suggests that with loss data for six years, a premium can be set that is both appropriate actuarially and gives the medical staff a meaningful incentive to conduct rigorous peer review. In contrast to the individual-based results in table 5, these estimates reflect past severity as well as claims frequency.

Based on the six-year variances and means shown in table 7, an assumed profit margin of three percent, and a probability of insolvency of 2.5%, we computed required initial capitalization (w) by hospital size category. Capitalization values by bedsize category are (with amounts per physician staff member in parentheses): $260,000 ($8,700) – under 100; $378,000 ($4,200) – 100 to 249; $1,059,000 ($5,800) – 250 to 399; and $990,000 ($3,600) – more than 400. Although the general pattern of w is plausible, the value for the 250–399 bedsize category is unusually high. This reflects a high variance that depends on the physicians that happened to be drawn to form the hospital medical staffs. In the 'real world,' physicians are not assigned to hospitals randomly. Hospital-based experience rating would probably make it more likely than now for physicians of like quality to be attracted to each other.

Within the smallest bedsize category, the hospital with the least favorable loss experience should establish a reserve of $660,000 ($22,000) and $15,000 ($500) for the hospital with the best experience. Clearly the size of the medical staff is only one determinant of required capitalization. Hospital medical staffs with 'risky' doctors would have to put up substantial amounts of money 'up front'.

To examine the impact of varying the observational period on the estimated capitalization amounts, we computed w using a five-year period (eliminating the 1975 observations). Estimated w was higher for all bedsize categories; not surprisingly, the increases were greater for the smaller hospital bedsize categories – 13% for the two smallest categories. For the 400+ bedsize category, the increase was only 0.2%. Lowering the probability of insolvency (ε) from 2.5 to 1.0% raises required capitalization by 19%. A higher profit margin than 3.0% would have reduced required capitalization and conversely for a lower profit margin.

5. Further discussion and conclusions

Without knowing a physician's claims record, it is possible to classify physicians into a few risk classes. There is no reason to limit the classifying variables to the few that have been employed by medical malpractice insurers to date. However, some of the classifying variables are plausible proxies for some ummeasured factors. Insurers are likely to have some information on physicians that is not available on public use data bases (or they should collect such information). Examples are prior disciplinary actions by licensing boards and/or hospitals and service and patient mix (including performance of high malpractice risk procedures, e.g., hair transplant, obesity suction, IV anesthesia, acupuncture). With such data, some of the proxy variables that have statistically significant impacts in our study may no longer attain significance.

There is clearly considerable within-group variation in loss for the risk classes we defined. There is only so much that the insurer can know about the doctor at the time he is insured. Refinements can be made by combining experience-rated premiums with those based on nonexperience-rated risk classes. Just conducting individual-based experience rating without prior risk classification results in discount-surcharge schedules that are probably too sensitive to changes in claims experience. A very steep schedule may expose physicians to excessive classification risk with the result that they will set the care level too high to avoid higher premiums later [Palfrey and Spatt (1985)]. Furthermore, the insured may fail to report claims to the insurer because this would change his risk classification [Hosios and Peters (1989), Lemaire (1985)]. Nonreporting is common in countries with substantial automobile insurance experience rating [Lemaire (1988)]. Insurers want to hear about cases early because they may be able to prevent more litigation and higher awards downstream.

Basing premiums on the experience of the hospital's medical staff seems feasible, at least from a statistical viewpoint. Judged in terms of credibility and desired initial capitalization, even medical staffs as small as 30 should be able to form their own groups. Setting the probability of bankruptcy for a six-year period at 2.5%, each physician on the staff of a hospital with less than 100 beds would be asked to make an initial capital contribution of 4.2 times the annual premium. This would be an attractive option for medical staffs with relatively favorable experience, such as Hospital No. 4 in table 7. The capital contribution per doctor in this hospital would be only \$500. A hospital initiating more rigorous peer review and/or more selective hospital privileges may reduce initial capital contributions below those based on historical experience (both because the mean *and* variance may be expected to decline).

Our simulations help one to think about the option of rating at the level of the hospital's medical staff. Our method does not capture the behavioral responses that are likely to occur – more careful staff selection and more rigorous peer review. To learn about such responses will require analysis of data from actual experience.

Although 'hard' empirical evidence is lacking, it does not appear that hospital-based peer review has been excessively rigorous [Blumstein and Sloan (1988)]. Thus, an added incentive may well be appropriate. Risk classification error should be less of a problem with in-depth reviews performed at the level of the hospital staff. Certainly pertinent information should be more complete on average. The peer reviewers should be well-positioned to distinguish good physicians with bad luck from 'bad apples'. Although we have focused attention on the hospital, essentially the same analysis is applicable to groups of physicians formed for other reasons, such as Health Maintenance Organizations.[27]

[27]Stevens (1981) provides a general discussion of medical malpractice in HMOs, emphasizing implications of contract and arbitration.

References

Abraham, K.S., 1986, Distributing risk: Insurance, legal theory, and public policy (Yale University Press, New Haven, CT).

Amemiya, T., 1973, Regression analysis when the variance of the dependent variable is proportional to the square of its expectation, Journal of the American Statistical Association 68, no. 344, 928–934.

Bawcutt, P.A., 1982, Captive insurance companies: Establishment, operation, and management (Dow-Jones Irwin, Homewood, IL).

Blumstein, J.F. and F.A. Sloan, 1988, Antitrust and hospital peer review, Law and Contemporary Problems 51, no. 2, 7–92.

Boyer, M. and G. Dionne, 1989, An empirical analysis of moral hazard and experience rating, Review of Economics and Statistics 71, no. 1, 128–134.

Bühlman, H., 1967, Introductory report: Experience rating and credibility, ASTIN Bulletin 4, 199–207.

Chang, L. and W.B. Fairley, 1978, An estimation model for multivariate insurance rate classification, in: Massachusetts Division of Insurance, Automobile insurance risk classification: Equity and accuracy (Division of Insurance, Commonwealth of Massachusetts, Boston, MA) 25–55.

Cooil, B., 1991, Using medical malpractice data to predict the frequency of claims: A study of Poisson process models with random effects, Journal of the American Statistical Association, forthcoming.

Cummins, J.D., B.D. Smith, R.N. Vance and J.L. VanDerhei, 1983, Risk classification in life insurance (Kluwer-Nijhoff, Boston, MA).

Danzon, P.M., 1985, Medical malpractice: Theory, evidence, and public policy (Harvard University Press, Cambridge, MA).

Ellis, R.P., C.L. Gallup and T.G. McGuire, 1990, Should medical professional liability insurance be experience rated?, Journal of Risk and Insurance 57, no. 1, 66–78.

Ferreira, J., Jr., 1978, Identifying equitable insurance premiums for risk classes: An alternative to the classical approach, in: Massachusetts Division of Insurance, Automobile insurance risk classification: Equity and accuracy (Division of Insurance, Commonwealth of Massachusetts, Boston, MA) 74–120.

Gourman, G., 1977, The Gourman report – A rating of American and international universities (National Education Standards, Los Angeles, CA).

Hewitt, C.C., Jr., 1970, Credibility for severity, Casualty Actuarial Society Proceeding 57, no. 107, 148–171.

Hewitt, C.C. and B. Lefkowitz, 1979, Methods for fitting distributions to insurance loss data, Proceedings of the Casualty Actuarial Society 66, no. 125, part 1, 139–160.

Hosios, A.J. and M. Peters, 1989, Repeated insurance contracts with adverse selection and limited commitment, Quarterly Journal of Economics 104, no. 2, 229–253.

Insurance Services Office, 1980, Report of the credibility subcommittee: Development and testing of empirical Bayes credibility procedures for classification ratemaking, Unpublished.

Landes, E.M., 1982, Insurance, liability, and accidents: A theoretical and empirical investigation of the effect of no-fault accidents, Journal of Legal Studies 25, 49–65.

Lemaire, J., 1985, Automobile insurance: Actuarial models (Kluwer-Nijhoff, Boston, MA).

Lemaire, J., 1988, A comparative analysis of most European and Japanese bonus-malus systems, Journal of Risk and Insurance 55, no. 4, 660–681.

Little, R.J.A., 1985, A note about models for selectivity bias, Econometrica 53, no. 6, 1469–1474.

Manning, W.G., N. Duan and W.H. Rogerts, 1989, Monte Carlo evidence on the choice between sample selection and two-part models, Journal of Health Economics 35, no. 1, 59–82.

Moore, M.J. and W.K. Viscusi, 1989, Promoting safety through workers' compensation: The efficacy and net wage costs of injury insurance, The Rand Journal of Economics 20, no. 4, 499–515.

Morris, C.N. and J.E. Rolph, 1981, Introduction to data analysis and statistical inference (Prentice-Hall, Englewood Cliffs, NJ).

Morris, C. and L. van Slyke, 1978, Empirical Bayes methods for pricing insurance classes, Proceedings of the Business and Economics Statistics Section (American Statistical Association), 579–582.

New York State Department of Health, 1988, Monitoring health care quality: Malpractice, misconduct, quality assurance (New York, NY).

Nye, B.F. and A.E. Hofflander, 1988, Experience rating in medical professional liability insurance, Journal of Risk and Insurance 55, no. 1, 150–157.

Nye, D.J., D.G. Gifford, B.L. Webb and M.A. Dewar, 1988, The causes of the medical malpractice crisis: An analysis of claims data and insurance company finances, Georgetown Law Journal 76, no. 4, 1495–1561.

Palfrey, T.R. and C.S. Spatt, 1985, Repeated insurance contracts and learning, The Rand Journal of Economics 16, no. 3, 356–367.

Rolph, J.E., 1981, Some statistical evidence on merit rating in medical malpractice insurance, Journal of Risk and Insurance 48, no. 2, 247–260.

Rothschild, M. and J. Stiglitz, 1976, Equilibrium in competitive insurance markets: An essay on the economics of imperfect information, Quarterly Journal of Economics 90, no. 4, 629–649.

Shayer, N., 1978, Driver classification in automobile insurance, in: Massachusetts Division of Insurance, Automobile insurance risk classification: Equity and accuracy (Division of Insurance, Commonwealth of Massachusetts, Boston, MA) 1–24.

Sloan, F.A., 1990, Experience rating: Does it make sense for medical malpractice?, American Economic Review 80, no. 2, 128–133.

Sloan, F.A., R.R. Bovbjerg and P.B. Githens, 1991, Insuring medical malpractice (Oxford University Press, New York, NY) forthcoming.

Sloan, F.A. and C.R. Hsieh, 1990, Variability in medical malpractice payments: Is the compensation fair?, Law and Society Review, forthcoming.

Sloan, F.A., P.M. Mergenhagen, W.B. Burfield, R.R. Bovbjerg and M. Hassan, 1989, Medical malpractice experience of physicians: Predictable or haphazard?, Journal of the American Medical Association 262, no. 23, 3291–3297.

Stevens, C.M., 1981, Medical malpractice: Some implications of contract and arbitration in HMOs, Milbank Memorial Fund Quarterly/Health and Society 59, no. 1, 59–88.

Taylor, G.C. and R. Buchanan, 1988, The management of solvency, in: J.D. Cummins and R.A. Derrig, eds., Classical insurance solvency theory (Kluwer Academic Publishers, Boston, MA) 49–155.

Topel, R.H., 1984, Experience rating of unemployment insurance and the incidence of unemployment, Journal of Law and Economics 27, 61–90.

Venter, G., 1983, Transformed beta and gamma distributions and aggregate losses, Proceedings of the Casualty Actuarial Society 70, no. 133, 156–193.

Victor, R.B., 1985, Experience rating and workplace safety, in: J.D. Worrall and D. Appel, eds., Workers' compensation benefits – Adequacy, equity, and efficiency (Cornell University, New York, NY).

U.S. General Accounting Office, 1986, Medical malpractice insurance costs increased but varied among physicians and hospitals (Government Printing Office, Washington, D.C.)

Weisberg, H.I., and T.J. Tomberlin, 1982, A statistical perspective on actuarial methods for estimating pure premiums from cross-classified data, Journal of Risk and Insurance 44, no. 4, 539–563.

Worrall, J.D. and R.J. Butler, 1988, Experience rating matters, in: P.S. Borba and D. Appel (eds.), Workers' compensation insurance pricing (Kluwer, Boston, MA) 81–94.