Malpractice Claims Data as a Quality Improvement Tool: II. Is Targeting Effective?

John E. Rolph, Richard L. Kravitz, Kimberly McGuigan
The research described in this report was supported by the Robert Wood Johnson Foundation under Grant No. 14556.

This Note contains an offprint of RAND research originally published in a journal or book. The text is reproduced here, with permission of the original publisher.

The RAND Publication Series: The Report is the principal publication documenting and transmitting RAND's major research findings and final research results. The RAND Note reports other outputs of sponsored research for general distribution. Publications of RAND do not necessarily reflect the opinions or policies of the sponsors of RAND research.

Published 1991 by RAND
1700 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138
Malpractice Claims Data as a Quality Improvement Tool: II. Is Targeting Effective?

John E. Rolph, Richard L. Kravitz, Kimberly McGuigan

Supported by the Robert Wood Johnson Foundation
II. Is Targeting Effective?

John E. Rolph, PhD; Richard L. Kravitz, MD, MSPH; Kimberly McGuigan, MS

Objective.—To evaluate the usefulness of malpractice claims data for identifying (1) physicians who are prone to negligent errors and (2) physician and hospital characteristics associated with particular kinds of errors.

Design.—Retrospective review of physician malpractice claim records.

Setting.—Large New Jersey physician malpractice insurer.

Participants.—Physicians practicing obstetrics and gynecology, general surgery, anesthesiology, or radiology and covered by the insurance carrier for any portion of 1977 through 1989.

Main Outcome Measures.—Claims were classified into 11 clinical error categories comprising three broad groups: patient management problems, technical performance problems, and staff coordination problems. Outcomes were expressed as per-physician frequency of claims due to negligence and proportion of claims associated with various types of errors.

Results.—Using 5 years of claims history to predict long-term claims proneness was more accurate than chance alone by 57% in obstetrics and gynecology, 33% in general surgery, 11% in anesthesiology, and 15% in radiology. Cross-validated recursive partitioning showed that among physician characteristics, only specialty was predictive of physician error profiles. For physician claims arising in acute care hospitals, hospital size and location in addition to hospital services discriminated among different error profiles; the cross-validated accuracy of this method was 69% compared with 22% accuracy achieved by random prediction.

Conclusion.—Use of physicians’ malpractice claims histories to target individuals for education or sanctions is problematic because of the only modest predictive power of such claims histories.

TWO theories dominate current thinking in medical quality improvement, each supported by critical assumptions about the sources of poor-quality care. The theory of traditional quality control as applied to medical care holds that bad care arises from the persistent practices of a limited number of individuals who, once identified, could be educated or otherwise motivated to perform better.

The theory of continuous quality improvement, on the other hand, emphasizes the careful examination, statistical analysis, and adjustment of clinical processes.

From the standpoint of efficiency, the choice of theories depends in part on the way in which clinical errors are distributed among a population of health care providers. If a large proportion of errors are concentrated among relatively few practitioners, the traditional quality control approach that targets error-prone clinicians for specific attention has high potential payoff. If clinical errors are more diffusely distributed, then targeting individual physicians may be a less efficient strategy than investigating the clinical processes in which many physicians are involved.

In a companion piece (this issue of JAMA), we described the types of negligent errors occurring among physicians in four specialties in New Jersey. In this article, we examine the concentration of clinically consequential errors (as represented by malpractice claims arising from negligence) among New Jersey physicians and explore the implications for insurance premium setting, risk management, and quality improvement. In addition, we evaluate the potential of malpractice claims data to identify specific groups of providers for interventions aimed at reducing specific types of medical negligence.

In so doing, we addressed three research questions. First, to what extent could the malpractice claims histories identify physicians with a propensity to incur negligence claims in the future? Second, what physician characteristics, if any, were associated with making particular types of clinical errors? Third, what hospital characteristics, if any, were associated with making particular types of clinical errors?
The answers to these questions bear on the more general issue of when quality assurance efforts should be directed at individuals and when they should be directed at clinical processes. If relatively few practitioners accounted for a large proportion of negligent care and if they could be successfully identified, then they could be appropriately educated or sanctioned. More specific targeting might be achieved if physicians with particular characteristics were prone to making particular kinds of clinical errors (e.g., surgeons prone to technical performance errors but not medication errors). Such physicians might benefit from highly tailored interventions (e.g., educational programs designed to strengthen their knowledge of biliary tract anatomy but not the use of perioperative antibiotics). On the other hand, if negligence were more evenly distributed throughout the physician population, targeting of individuals might not be efficient, fair, or even possible.

**METHODS**

The creation of the analytic files for this study required three steps: (1) identifying medically negligent injuries, (2) creating a taxonomy of clinical error types, and (3) linking information on physicians and hospitals to malpractice claims records.

**Selection of Claims**

Detailed information on New Jersey malpractice claims against physicians was obtained from the database of the Medical Inter-Insurance Exchange of New Jersey (the "Exchange"), a physician-owned, medical society-sponsored malpractice insurer covering approximately 70% of the physician malpractice market statewide. Claims were selected from among all claims filed between 1977 and 1989 against physicians practicing in the specialties of obstetrics and gynecology, general surgery, anesthesiology, and radiology. Because previous analyses of these data showed virtually no relationship between individual physician claims frequency and size of indemnity payment, we did not use claims severity in this analysis. For the analysis of claims-proneness, as well as the analysis of provider type-error type association, we selected claims on which payment had been made (excluding chipped teeth claims against anesthesiologists) or for which peer review had determined negligence even in the absence of payment, calling them negligence claims. Claims not meeting these criteria were excluded in part because of prior work showing no-pay claims frequency to be only weakly correlated with paid claims frequency.

**Creation of a Clinical Error Taxonomy**

The Exchange maintains a detailed risk prevention file on every claim brought against its policyholders. This file is developed and maintained by experienced risk prevention analysts; it contains objective data and judgments by the analysts as to the types of clinical errors leading to injury.

We combined detailed judgments (termed "behavior errors") into 11 clinically coherent categories called "aggregated misadventures" (Figure). We further combined the 11 categories into three major problem types: patient management problems (problems with recognition, diagnosis, and judgment); technical performance problems (technical lapses during operative or nonoperative procedures); and medical and nursing staff coordination problems (resulting from the interaction of at least two health care providers). The scheme had acceptable face validity and interrater reliability.

**Linkage of Provider Characteristics and Malpractice Claims Data**

Information on the self-reported characteristics of physician policyholders and the names of the hospitals where they maintained staff privileges were obtained from the Exchange. Physician variables (as of January 1987) included age, gender, specialty, whether board certified, quality of residency (1 = flagship teaching hospital, 0 = other), number of office patients per week, hours worked per week, volume of surgery, and performance of various high-risk procedures.

A comprehensive database for New Jersey hospitals was linked to the Exchange's database on physician characteristics and claims, permitting analysis of the relationship between physician and hospital characteristics and the types of errors leading to claims. Available hospital variables included size and type of hospital, teaching status, percentage of staff who were board-certified, geographic setting (urban or rural), hospital financial status (e.g., Medicare billings, revenues), payer mix.
Statistical Analysis

Predicting Claims-Proneness.—A physician's claims-proneness is defined as his or her unobservable expected annual rate of incurring negligence claims, denoted by \( g \). That physician's actual number of negligence claims per exposure-year will follow a Poisson distribution with a mean of \( g \). If all physicians in a given specialty had precisely the same value of \( g \), say \( g' \), then the number of negligence claims per physician aggregated across the specialty would also follow a Poisson distribution with mean \( g' \). However, if the observed distribution of negligence claims over time was more spread out than a Poisson distribution (i.e., had longer tails), then some physicians in the group would be relatively more claims-prone (have a larger \( g \)) and some less so. This is the case in each of the four specialties examined in this report (although we used a .10 significance level cutoff to include anesthesia in this statement).

To capture the variability in individual values of \( g \) within each specialty, we assumed a gamma distribution as the functional form. Fitting a gamma mixture of the individual Poisson distributions of claim counts produces a negative binomial distribution of individual claim counts. We fit this model separately by specialty to the complete database by the maximum likelihood principle using the software package GLIM.

Using the gamma distribution of \( g \) as estimated from all the data, we calculated the posterior distribution based on the most recent 5 years of negligence claims. This allowed us to estimate how well 5 years of data could predict each physician's (unobserved) value of \( g \). To measure the accuracy of how well a physician's actual number of negligence claims predicted his or her claims-proneness, we calculated the agreement of the two sets of rankings within each of the four specialties. The first ranking was formulated according to the estimated (but unknown) values of \( g \), while the second ranking was by the observed number of claims per physician for a 5-year period. To compare the agreement of the two rankings, we calculated how well the observed 5 years of claims experience identified which physicians were in the top 25% and in the top 10% of this rank ordering of estimated \( g \). The relative improvement over chance (RIOC) measures how accurately the prediction of who is more prone to incur claims compares to chance prediction.

In brief, we asked whether all physicians in a specialty had the same value of \( g \). If not, we estimated the distribution of \( g \) within the specialty being analyzed. Based on this estimated distribution, we computed the proportion of claims-prone (i.e., "high \( g' \)) physicians who would be correctly identified using actual past claims and compared this with the percentage that one would expect to predict correctly based on chance alone.

We chose to use a physician's \( g \) rate as our classification criterion rather than the actual number of negligence claims against the physicians in the succeeding, say, 5 years to avoid introducing additional classification error. The additional error would arise because of the substantial year-to-year variation in an individual's claims rates. The use of a statistical model to generate the frequency distribution of actual claims-proneness for a physician population gets around this problem.

Recursive Partitioning.—To investigate the relationship between clinical error type and characteristics of the physician against whom claims were filed, we used a computerized recursive partitioning algorithm (CART) to construct negligence claim classification trees. Tree-based models have been employed increasingly in a variety of medical studies. Characteristics of tree-based models advantageous to this study include (1) their ability to capture complex interactions between predictor characteristics in predicting clinical error type (as compared with the prespecified multiplicative form of interactions in linear models, such as discriminant analysis and logistic regression); (2) easier interpretation of the mix of continuous and categorical predictors; and (3) ease of cross-validation. The algorithm's first step is to identify the single physician or hospital characteristic that most strongly differentiates across types of clinical errors; it then divides the claims population into two groups defined by the presence or absence of this characteristic, each being more homogeneous in its mix of clinical error types than the overall population. Each subgroup is then divided using the variable that is most predictive of the error type distribution for that subgroup. The process continues iteratively until the last candidate subgroup for a split either consists of only one error type or is too small to be split further. The CART then uses 10-fold cross-validation to prune this full tree back to a smaller, statistically stable classification tree.
Table 1.—Percentage of Claims-Prone Physicians (Top Quartile) Correctly Classified Using Past Claims

<table>
<thead>
<tr>
<th>Specialty</th>
<th>Significance Probability for Claims Proneness*</th>
<th>Percentage of Actual Correct Prediction</th>
<th>Percentage of Correct Prediction at Random</th>
<th>Percentage of Relative Improvement Over Chance†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anesthesiology</td>
<td>0.084</td>
<td>66</td>
<td>63</td>
<td>11</td>
</tr>
<tr>
<td>General surgery</td>
<td>0.025</td>
<td>76</td>
<td>63</td>
<td>33</td>
</tr>
<tr>
<td>Obstetrics and gynecology</td>
<td>&lt;.001</td>
<td>84</td>
<td>63</td>
<td>57</td>
</tr>
<tr>
<td>Radiology</td>
<td>0.029</td>
<td>68</td>
<td>63</td>
<td>15</td>
</tr>
</tbody>
</table>

*The Lawless S test statistic against extra-Poisson variation and has a χ² distribution with 1 df: the corresponding P values are given.
1A γaamma Poisson distribution was fit to the number of paid claims per physician. The γaramma shape and scale parameters are (0.21, 568), (0.67, 772), (1.16, 953), and (0.29, 604), for the anesthesiology, general surgery, obstetrics and gynecology, and radiology specialties, respectively. The proportion correct is computed by separately summarizing across the gamma posterior distributions corresponding to physicians above and below the upper quartile of the observed paid claims distribution to get the true-positive and true-negative proportions.
†Relative improvement over chance (p₀−pc), where pc is the actual proportion correct and p₀ is the proportion correct classification with random prediction.

Table 2.—Anesthesiology Errors by Hospital Type

<table>
<thead>
<tr>
<th>Hospital Type</th>
<th>Patient Management Errors No. (%)</th>
<th>Technical Performance Errors No. (%)</th>
<th>Coordination Errors No. (%)</th>
<th>Claims by Hospital Type No. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rural</td>
<td>4 (50)</td>
<td>3 (38)</td>
<td>1 (12)</td>
<td>8 (100)</td>
</tr>
<tr>
<td>Small, not rural</td>
<td>19 (58)</td>
<td>10 (30)</td>
<td>4 (12)</td>
<td>33 (100)</td>
</tr>
<tr>
<td>Large, inner city</td>
<td>17 (49)</td>
<td>18 (51)</td>
<td>0 (0)</td>
<td>35 (100)</td>
</tr>
<tr>
<td>Large, teaching; not rural or inner city</td>
<td>27 (44)</td>
<td>29 (48)</td>
<td>5 (8)</td>
<td>61 (100)</td>
</tr>
<tr>
<td>Large; not rural, inner city, or teaching</td>
<td>17 (47)</td>
<td>16 (44)</td>
<td>3 (6)</td>
<td>36 (100)</td>
</tr>
<tr>
<td>Overall</td>
<td>84 (49)</td>
<td>76 (44)</td>
<td>13 (8)</td>
<td>173 (100)</td>
</tr>
</tbody>
</table>

*Percentages are type of error for each hospital type and hence sum to 100 across the rows (except for rounding).
1The P value for the χ² test is .056.
†Rural indicates hospitals in rural locales; small, admissions <7500 annually; large, inner city, >7500 admissions annually with Medicaid admissions at least 20%.
§Number of negligence claims against physicians in anesthesiology in each type of hospital.

characteristic (including specialty) and the proportion of negligence claims associated with management, performance, or coordination problems. Applying the same classification procedure to the 11 aggregated misadventure types resulted in a cross-validated tree with two terminal regions (radiologists and all others), confirming the relative propensity of radiologists to incur claims associated with diagnostic errors.7

Using Hospital Characteristics to Predict Physician Error Profiles

Of 1871 physician claims in our database, 83% occurred in acute care hospitals. Within each of the four specialties, no individual hospital characteristic (see "Methods" section) was significantly associated with any of the three broad error types (P>.10) (data not shown in tabular form). However, when we divided New Jersey hospitals into five categories based on size, geographic area, teaching status, and percentage of Medicaid admissions, there was a lower proportion of anesthesiology performance errors in small nonrural hospitals (30% compared with 44% on average) and a correspondingly larger percentage of patient management errors (58% vs 49%) (P=.056; Table 2). There were no statistically significant pairwise associations in the other three specialties (P>.10).

In the multivariate analysis we used classification trees to predict outcome for the three broad error types, as well as the 11 aggregated misadventure types. In predicting the three broad error types, no cross-validated tree could be produced, indicating the absence of stable predictors. For the 11 aggregated misadventure categories, the cross-validated tree had seven terminal regions (Figure). Radiology claims and then anesthesia claims were the initial "splits," with radiology claims dominated by diagnostic errors (74%), and anesthesia claims consisting largely of unintentionaliatrogenic injuries (44%) and decision errors (20%). The CART successfully split the remaining obstetrics and gynecology and general surgery claims into five clusters using five variables: ratio of hospital staff to admissions, percentage of self-paying patients, percentage of Medicare patients, number of beds, and a combination of size and geographic location (small or rural vs other). The largest terminal region (No. 5, Figure) consisting of surgery and obstetrics and gynecology claims from hospitals with high staff-to-admission ratios, accounted for half of all claims (n=723). The Figure displays the two most prevalent misadventure categories for each of the seven terminal regions of the tree. The cross-validated classification accuracy of the tree was 69%, as compared with 22% by random prediction and 33% by predicting all outcomes to be the most prevalent category.

In summary, physician error profiles displayed statistically stable variation by hospital service (specialty) and to a lesser degree, by selected hospital characteristics (size, location, staffing, and patient mix).

COMMENT

From a policy perspective, this study addresses two related questions. First, can we use malpractice claims experience to predict which physicians are claims-prone? Second, can we use the claims experience of physicians to predict who is going to incur relatively more claims of a particular type? The results were somewhat disappointing. Although some New Jersey physicians in the four specialties examined were statistically more prone to incur claims than others, attempts to identify them prospectively based on available past claims data were only moderately successful. Furthermore, aside from specialty, no physician characteristics were associated with particular error profiles. While certain hospital characteristics were associated with particular physician error profiles, only a small proportion of hospitals were identified as distinctive.

Identifying Error-Prone Physicians

Previous research has clearly established that the risk of incurring paid malpractice claims varies between specialties and between individuals within specialties.13,234 The results of the current study indicate that past claims experience is only modestly predictive of intrinsic claims-proneness. Although physicians incurring large numbers of negligence claims in the past are more likely, on average, to incur large numbers in the future, predictions about individuals based on past claims experience are probably not accurate enough to identify most claims-prone physicians or to allow reliable judgments about an individual's propensity to practice negligently in the future.

In this light, other means of identifying negligence-prone physicians might be considered. One method used by some carriers is to review a sample of office records obtained from practitioners deemed responsible for especially egregious mistakes or from physicians who practice in particularly high-risk specialties.28 The basic method is well

2096 JAMA, October 16, 1991—Vo1266, No. 15 Malpractice Claims Data—Rolph et al
established but its application to physicians with poor malpractice records is unproved.

One reason for the difficulty in identifying error-prone physicians using malpractice claims data is that malpractice claims (both meritorious and otherwise) are relatively rare events (mean rate in our database, 0.15 total claims per physician per year). When event rates are higher, as in complications from carotid endarterectomy (mean rate, 11.3%), past experience may be a better predictor of future outcomes.34

**Identifying Physicians With Distinct Error Profiles**

When we began this study, we had hoped to identify types of physicians (eg, male gynecologists older than 60 years who practiced less than 35 hours per week; female anesthesiologists younger than 45 years who performed a large number of nerve-block procedures) who were more likely to incur some types of negligence claims than others. Identifying the kinds of physicians prone to making certain kinds of errors would have allowed us to suggest specific deterrent or educational strategies targeted at particular groups of physicians. The design of such strategies requires both systematic differences in profiles of error types across providers and a practical way to prospectively identify groups of providers with particular types of error profiles.

Among all physician characteristics we examined, only specialty was consistently associated with error type. Thus, reducing the incidence of particular types of errors by applying selective deterrence policies is problematic because of the lack of a reliable way to identify target groups of physicians. These results are consistent with recent studies showing structural variables to be poor predictors of quality of care.4,5

**Identifying Hospitals With Distinct Profiles of Physician Errors**

As in the analysis of physician characteristics, hospital service (specialty) was the most important predictor of physician error pattern, followed by the ratio of hospital staff to admissions (Fig). Other hospital variables had some predictive value, but the results were applicable only to a small proportion of negligence claims. For example, 50% of claims in obstetrics and gynecology and general surgery arising in hospitals with a low staff-to-admissions ratio and a low percentage of self-pay patients (terminal region 1, Figure) were attributable to problems in communication between providers (compared with an overall average of 2%), but these hospitals accounted for less than 1% of negligence claims in the database. Nevertheless, insurance carriers considering intervention programs to help hospitals with interdepartmental communication might consider piloting them in hospitals with these characteristics.

**Implications for Quality Improvement**

In this analysis of New Jersey physicians, we found little evidence that negligence claims were sufficiently concentrated, either in number or in kind, to permit negligence reduction strategies targeted at individuals. Our results do not contradict those of Sloan et alii but rather emphasize the difference between knowing that some practitioners must be more prone to incur claims than others (because the observed tail is much longer than expected) and being able to identify who they are.

These findings suggest caution in the use of the newly established National Practitioner Data Bank, which was founded on the premise that paid malpractice claims or adverse National Board of Examiners reports can be used prospectively to identify practitioners who are likely to have future problems. While the data bank may be useful in identifying some extremely problematic physicians, it may have even greater utility as a national repository of information on consequential clinical errors, thereby supporting efforts to improve clinical processes in the spirit of continuous quality improvement.

This study was funded by grants from The Robert Wood Johnson Foundation, Princeton, NJ, and the RAND Corporation, Santa Monica, CA.

Peter Sweetland (deceased), William Reilly, Adam Wilczek, and Howard Weiss of the Medical Inter-Insurance Exchange of New Jersey, Lawrenceville, NJ, shared their claims database and graciously responded to our many queries in the course of the study. Elizabeth Lewis, MS, and Albert Mello contributed to the preparation of the analysis files. Glenn Meichin, PhD, and Joyce Mann, PhD, provided us with their database on New Jersey hospitals. Helen Blumen, MD, validated the clinical error taxonomy. Helen Blumen, MD, and Robert Brook, MD, ScD, Joel Cantor, PhD, Allan Cohen, PhD, Emmett Keeler, PhD, Robert Klitgaard, PhD, Don Harper Mills, MD, JD, William Reilly, Elizabeth Rolph, MA, Susan Dчит, MD, and Adam Wilczek provided useful suggestions but are in no way responsible for opinions and findings. Dee Huthinson was responsible for typing the various revisions of this report.

**References**
