JUDICIAL PROTECTION OF MANAGED CARE CONSUMERS: AN EMPIRICAL STUDY OF INSURANCE COVERAGE DISPUTES†

*Mark A. Hall**
*Teresa Rust Smith***
*Michelle Naughton****
*Andrea Ebbers*****

Most discussions of consumer protection in managed care settings, such as health maintenance organizations (HMOs), focus on state or federal regulatory oversight.[1] This oversight usually takes the form of plan-level due process rights, such as requiring notice of adverse decisions and opportunity to initiate grievance and appeal procedures. As important as these regulatory protections are, they overlook an equally important and more obvious source of protection: the courts. Patients' relationships with managed care organizations are grounded in contract or contract-like principles derived from entitlement statutes. This legal regime creates the opportunity to seek judicial resolution of grievances against HMOs, utilization reviewers, and other managed care entities.

Under private-sector insurance, the right to sue exists by virtue of the insurance contract's promise of a certain level of benefits. Under public-sector insurance, such as Medicare and Medicaid, the same right to sue for covered benefits exists by virtue of the defined-benefit structure of these programs. They promise a defined level of benefits in terms almost identical to those that prevail in private insurance contracts. Both sources of insurance typically

---

† This Article was delivered at the Symposium on Consumer Protection in Managed Care, on November 17, 1995, at Seton Hall University School of Law. The Agency for Health Care Policy and Research, grant no. 1 R01 HS08681-01 funded this study.

* Professor of Law and Public Health, Wake Forest University, Bowman Gray School of Medicine. This paper is based on a talk I presented at the Seton Hall Law School while serving as the Merck Distinguished Professor of Health Law. I am grateful for this support, and for the helpful comments of Gerard Anderson and Stewart Schwab.

** Research Associate, Bowman Gray School of Medicine; Visiting Professor in Sociology, Wake Forest University.

*** Assistant Professor, Bowman Gray School of Medicine.

**** Research Specialist, Bowman Gray School of Medicine.

1 *See, e.g.*, Susan J. Stayn, Note, *Securing Access to Care in Health Maintenance Organizations: Toward a Unified Model of Grievance and Appeal Procedures*, 94 COLUM. L. REV. 1674 (1994).

define covered benefits in terms of "medical necessity." Described in a nontechnical fashion, medically necessary means medically appropriate. It excludes experimental care, nonstandard treatments, treatment without any known benefit, and treatment such as cosmetic surgery not intended to correct or relieve a medical condition. Many issues of consumer protection relate directly to how these contractual or statutory entitlement concepts are interpreted and applied in practice. Therefore, how courts enforce these entitlements is central to the level of consumer protection that the law creates in managed care settings.

This Article analyzes judicial determinations of health insurance coverage disputes from an empirical perspective, using a method called "content analysis." This perspective contrasts with the conventional approach of simply reading and describing one's subjective interpretation of a selection of important decisions. Instead, we located all decisions of possible relevance and coded them comprehensively for a host of factors of potential interest, and then subjected this coded information to statistical analysis. This method produced a number of findings that are of central importance to gauging the level of consumer protection created by judicial review of insurance coverage disputes. We report on those findings here, first by describing in detail our methodology, and then by presenting the key quantitative results along with our interpretations of those results.

I. METHODS

We compiled a list of all health insurance coverage disputes involving issues of medical appropriateness that resulted in published federal and state court decisions from 1960 to June 1994. We included all forms of standard health insurance, both public and private, including Medicare and Medicaid, the Federal Employees Health Benefit Program and the Civilian Health and Medical Program of Uniformed Services (CHAMPUS), and self-insured employers. We included cases involving HMOs and managed care plans, although surprisingly we discovered only six of these that met our selection criteria. We excluded health insurance that was ancillary to another type of insurance such as automobile or worker's compensation. We limited our analysis to only published decisions, because unpublished decisions are not readily available and published decisions form the basis for determining legal precedents that influence future behavior and judicial decisions. When appeals were filed resulting in multiple published decisions

pertaining to a single case, only the highest decision relevant to the study was included.

For a case to be included in this study, the court's basis for decision had to turn on the appropriateness of the treatment rather than on a procedural technicality or some other legal defect in the plaintiff's case. We defined appropriateness disputes to encompass questions of whether the treatment was "medically necessary," "experimental," "investigational," "custodial," or other similar terms. Also included were cases that involved disputes where treatment-specific contractual exclusions were possibly related to the appropriateness or experimental status of the treatment, such as "no coverage for solid organ transplants" or "in vitro fertilization is not covered." Examples of exclusions that are clearly not related to medical appropriateness are those that merely define the type of insurance, such as exclusions of mental health or dental services. The study omits cases such as these.

A multi-step process was undertaken to identify all cases meeting these selection criteria. The primary search tools were the LEXIS and WESTLAW computer databases. In order to identify the best keywords to be used in a comprehensive search request, an initial list of keywords was used to generate a preliminary listing of cases. A review of the relevant cases from this pilot search suggested additional keywords and modifications in the search logic. In addition, fifty-three cases that had been identified as relevant through prior legal research and from published articles were examined for keywords and were used to supplement the initial list.

The following is the final search request used to search the WESTLAW database.

> DATE (AFT 1959) and HE, SY(((HEALTH HOSPITA! MEDICAL) /P (POLICY MEDICARE MEDICAID)) or (INSURANCE INSURED COVERAGE COVERED)) & HE, SY((NECESSARY APPROPRIAT! EXPERIMENTAL INVESTIGATIV!) /P (MEDICAL! PROCEDURE TREATMENT TRANSPLANT)) % HE, SY(DIVORCE ENVIRONMEN! MALPRACTICE "WORKERS COMPENSATION").[2]

This computer search identified a total of 3215 cases. A review of these cases determined that 1155 did not involve an insurance

[2] The first clause limits the search to cases after 1960. The search is then limited to the headnotes and syllabus. The first set of terms restricts the search to the medical arena, and the second and third sets of terms specify various types of insurance. The next two sets of terms specify the health insurance disputes of particular relevance to our inquiry, and the final set of terms seeks to exclude cases that are not relevant.

policy dispute; 940 did not involve a coverage issue; 757 did not involve a medical necessity dispute; and 160 involved either automobile or worker's compensation. After these exclusions, a total of 203 cases met the selection criteria.[3]

We then analyzed these cases using content analysis. Content analysis is a research technique widely used in the social sciences that allows the objective and quantitative study of qualitative data.[4] The first step in this analysis was the development of a code book based on a series of hypotheses concerning the potential factors that could influence judicial reasoning, drawn from the published literature.[5] The final code book contained thirty-seven questions which can be classified into the following general categories: case description (date, court, type of insurance, outcome); legal factors (source of law, contract language, level of review, stage of review); patient/treatment factors (patient demographics, type of treatment, seriousness of condition, results of treatment); evidence of treatment effectiveness (sources of assessments, numbers of expert witnesses, objections to assessments, process of review); and the court's analysis (whether various factors appeared to influence the court, the most important factors).

Once the code book was developed, two law students and a lawyer in medical school were trained to read and code the cases. Cases were assigned to one reader each, and coding reliability was assessed at periodic intervals by having approximately fifteen percent of the cases recoded by the other two readers. This is known as assessing "interrater reliability." Whenever interrater reliability fell below seventy percent on a specific question, the question was dropped from the analysis. In addition, a sample of the cases were

---

[3] For many court decisions it was not possible to determine from the written opinion an answer to one or more of our specific questions. As a result, the number of cases included in the analysis varies from question to question.

[4] T.F. CARNEY, CONTENT ANALYSIS: A TECHNIQUE FOR SYSTEMATIC INFERENCE FROM COMMUNICATIONS (1989).

[5] J.H. Ferguson et al., *Court-Ordered Reimbursement for Unproven Medical Technology: Circumventing Technology Assessment*, 269 J.A.M.A. 2116 (1993); Barbara A. Fisfis, Comment, *Who Should Rightfully Decide Whether a Medical Treatment Necessarily Incurred Should Be Excluded From Coverage Under a Health Insurance Policy Provision Which Excludes from Coverage "Experimental" Medical Treatments?*, 31 DUQ. L. REV. 777 (1993); Mark A. Hall & Gerard F. Anderson, *Health Insurers' Assessment of Medical Necessity*, 140 U. PA. L. REV. 1637 (1992); Angela R. Holder, *Funding Innovative Medical Treatment*, 57 ALBANY L. REV. 795 (1994); David C. Hsia, *Benefits Determination under Health Care Reform: Who Should Decide Coverage Policy?*, 15 J. LEGAL MED. 533 (1994); Richard S. Saver, Note, *Reimbursing New Technologies: Why are the Courts Judging Experimental Medicine?*, 44 STAN. L. REV. 1095 (1992).

abstracted by one of us (Hall), in order to monitor the validity (accuracy) of the students' coding responses.

These coding data were analyzed, first using bivariate analysis (comparing one factor to only one other) and then using multivariate analysis. Throughout the analysis, the primary dependent variable (the factor we were trying to predict) was the outcome of the case, defined as coverage upheld or coverage denied. Where a decision was not definitive, for instance, where it called for the case to be sent back to the trial court for additional findings, coders were asked to classify the decision as primarily favoring one party or the other, or as indeterminate. Only the determinate results were classified as coverage upheld (patient wins) or coverage denied (insurer wins).

II. PRINCIPAL FINDINGS

The principal findings will be divided into simple descriptive frequencies, single-variable effects on the outcome of the cases, and multivariate analysis of case outcomes.

A. Descriptive Frequencies

The first finding of interest is simply how few cases were located. Over a period of thirty-four years, we found only 203 published decisions of relevance, including federal and state jurisdictions, and both public and private health insurance. Partly, this was due to our criteria of relevance. Several hundred cases were excluded because they did not involve a question of medical appropriateness or they turned on purely jurisdictional issues. Nevertheless, the relative paucity of cases suggests that a great many coverage disputes exist that do not go through litigation, and the vast bulk of coverage decisions are never challenged at all. Our study looks only at the small tip of a large iceberg; we were unable to gain any knowledge of what occurs below this surface of reported decisions.

We can speculate on several reasons that coverage denials do not result in reported judicial decisions. Many of these decisions may be resolved amicably at the plan level. Decisions that are litigated may be settled before trial, or a trial decision may be accepted by the parties and not appealed, which in the state courts results in no reported decision. It is also likely, however, that patients do not pursue coverage disputes into court because of the

time and expense involved.[6] This is suggested by our finding that the median cost of treatment in reported decisions is between $10,000 and $50,000, and the mean cost is between approximately $50,000 and $100,000.[7] It is also suggested by the finding that cases took about 2.5 years on average to reach final decision. One quarter of the cases took four years or more, and fifteen percent took five years or more.

The trend in the number of cases per decade reveals a dramatically different picture, however. We found a geometric increase in the number of cases being litigated to a reported decision in recent years: 5 in the 1960s, 36 in the 1970s, 71 in the 1980s, and 200 in the 1990s.[8] Partly, this increase is an artifact of how judicial decisions are reported in federal versus state courts. The 1980s and 1990s saw a marked shift from state to federal courts as a consequence of ERISA preemption,[9] and in federal courts trial decisions are frequently reported. But this jurisdictional shift does not appear to fully account for the dramatic increase in the number of reported cases in recent years. Our distinct impression is that health insurance litigation is on the rise.

An equally dramatic finding is the near absence of decisions arising from a managed care setting. We attempted in several ways to identify cases that involve HMO insurance, but could locate only six of them. Several explanations for this finding are possible. First, it is not always possible to tell from the reported decision the precise structure of the insurance. Second, there is a time lag of several years between treatment denial and ultimate judicial decision, so our sample reflects mostly cases from the 1980s, when managed care was less prevalent. But this finding is also consistent with a more substantive explanation relevant to consumer protection.

Managed care settings such as HMOs are less likely to produce coverage disputes, even though they are more likely to deny treat-

---

6 *See* Margaret Gilhooley, *Broken Back: A Patient's Reflections on the Process of Medical Necessity Determinations*, 40 VILL. L. REV. 153 (1995) (providing a first-hand account from a patient's perspective of the difficulties in challenging medical necessity denial).

7 Precise figures are not available because cases were coded within specified ranges.

8 The 1990's figure was extrapolated from the first 4 1/2 years of the 1990s.

9 As most readers are well aware, ERISA is the federal statute that primarily governs employee pension plans, but also covers other employee benefits such as health insurance. *See* 29 U.S.C. § 1001 et seq. (1988 & Supp. V 1993) (ERISA). ERISA requires employees who challenge health benefits determinations to bring their suits in federal court. 29 U.S.C. § 1451(c).

ment, because coverage decisions are frequently made by treating physicians or by a medical director in the physician's practice group. Indemnity insurance requires a third-party medical director with the insurance company to overrule the treating physician's recommendation. HMOs and other forms of managed care, in contrast, use corporate and financial incentives to motivate physicians not to make treatment recommendations in the first place. This absence of patient knowledge that potentially beneficial care is being foregone could easily account for the lack of coverage disputes arising from managed care settings.

Although we identified few cases arising from an explicitly managed care setting, the cases we analyzed were broadly representative of many kinds of health insurance. As reflected in Table 1, our sample included approximately thirty percent Medicare and Medicaid cases, sixty percent private insurance (including employer self-insured), and ten percent other types such as government employees and union plans. Accordingly, these cases were divided fairly evenly among state contract law (thirty percent), state or federal statutory and regulatory law (thirty-five percent), and federal ERISA law (thirty-four percent). The jurisdictional split is further reflected in Table 2, which reports the division of cases among types of courts.

TABLE 1
CASES PER TYPE OF INSURANCE (PERCENT)

Medicare	17
Medicaid	13
Commercial	34
Blue Cross	18
Employer Self-insured	7
Taft-Hartley (Union) Plan	3
Federal and Military Employees	7

TABLE 2
CASES PER JURISDICTION (PERCENT)

State Supreme	9
State Appeals	26
State Trial	3
Federal Appeals	22
Federal Trial	39

B. Single-Factor Predictors of Outcome

Having described the cases we analyzed, we will now present the findings that reveal which factors of interest are significantly associated with the outcome of these cases. As discussed above, we defined the outcome according to whether the decision primarily favored the insurer by finding no coverage or favored the patient by finding the treatment in question was covered. Overall, fifty-seven percent of cases with some definitive outcome found for the patient. Naturally, there is no way to know whether this is over, under, or optimally protective. It is noteworthy, however, that this margin of victory persists across almost all of the major types of insurance (Medicare, Medicaid, Blue Cross, commercial, and self-insured employers) and across all major sources of law (state contract, state and federal statute). We will focus our attention on where this pattern is not borne out.

TABLE 3
OUTCOME BY TYPE OF INSURANCE

	Public	Private	Government Employees
Percent of Patient Wins	70	54	31

p=0.017

TABLE 4
OUTCOME BY SOURCE OF LAW

	State Contract	State Statute	Federal Statute	ERISA
Percent of Patient Wins	61	60	67	43

p=0.082

Cases brought under ERISA were found against patients at a rate significantly higher than under other sources of law.[10] We explored why this might be the case. One possibility is that ERISA applies a more lenient standard of review. We found no statistically significant difference, however, in who wins or loses according to

[10] This is reflected in notation p=0.082. This refers to the statistical probability (using the Chi-square test) that this observed difference could be the result of chance alone. When this probability falls to 0.05 or below, it meets standard conventions for statistical significance. When ERISA was compared with all other sources of law combined, the p value dropped to 0.009.

whether the reviewing court employs an arbitrary and capricious or a *de novo* standard of review. A related factor appeared to have a much stronger effect. Under ERISA, the standard of review is determined in part by the amount of deference the insurance contract assigns to the insurer. Contract language that allows the insurer to determine medical necessity "in its sole discretion" or language to similar effect was coded as assigning discretion to the insurer. In such cases, insurers prevailed much more often than when they were not given discretion. This is because language of this nature reverses the presumption that usually applies in insurance cases that ambiguities are to be construed in favor of the patient. Insurer discretion to interpret coverage language results in construing ambiguities to favor the insurer.

TABLE 5
EFFECT ON OUTCOME WHEN DISCRETION IS ASSIGNED TO INSURER

	Discretion Assigned	Discretion Not Assigned
Percent of Patient Wins	37	80

p=0.000

Our finding that discretion-assigning language strongly affects case outcome indicates that courts are sensitive to some aspects of how the insurance coverage language is drafted. Other findings suggest, however, that courts are not sensitive to the wording of the coverage language. Most surprising is the finding that the outcome of these cases did not differ significantly depending on whether the insurance contract, statute, or regulation used general coverage language such as "medically necessary" or "experimental," or instead specifically excluded the treatment in question.[11] This is because courts so frequently fault insurers for not making their coverage exclusions more explicit. When insurers attempt to correct this defect, however, they are criticized for making the exclusion too narrow or technical. A leading example is the attempt to deny coverage for autologous bone marrow transplant (ABMT) to treat breast cancer. Specific exclusions to this effect have been held not to apply to newer cases that use a modified technique known as "peripheral stem cell rescue," because the specific exclusion failed to anticipate and specify this newer approach.[12]

[11] *See* Table 6.

[12] *E.g.*, Frendreis v. Blue Cross Blue Shield of Michigan, 873 F. Supp. 1153 (N.D. Ill. 1995); Mattive v. Healthsource of Savannah, Inc., 893 F. Supp. 1559 (S.D. Ga.

TABLE 6
OUTCOME BY COVERAGE LANGUAGE

	General Coverage Language	Specific Exclusion
Percent of Patient Wins	57	48

p=0.412

Another factor that possibly explains why insurers prevail more often under ERISA is simply that they are in federal court. We found that patients did significantly worse in federal appeals courts than in either federal trial courts or state appeals courts. The reason for this association is not at all clear. ERISA results in cases being brought in federal court. However, so does Medicare. Further, ERISA also results in cases being brought to federal trial courts, yet patients prevailed more often under Medicare and in district court.

TABLE 7
OUTCOME BY COURT JURISDICTION

	Federal Appeals	State Appeals	Federal Trial
Percent of Patient Wins	39	62	63

p=0.027

Nevertheless, the federal appeals effect stands out clearly. One possible reason for this difference is simply that insurers were appellees in 66 percent of the federal appeals cases, but in only 45 percent of the state appeals cases. This alone could account for the federal appeals effect since, all other things being equal, appeals courts are more likely to affirm than to reverse a trial court opinion, owing to the deference given to findings of fact, evidentiary and procedural rulings, and other discretionary matters. We do not know why patients make up more of the appellants in federal court.

Another possible explanation for the different outcomes in different courts is that appeals courts are less swayed by sympathy for the plight of seriously ill patients. Unlike trial courts, they do not see the patient in person because they do not receive live testi-

1995); Wilson v. Office of Civilian Health and Medical Program of Uniformed Services (CHAMPUS), 866 F. Supp. 931 (E.D. Va. 1994).

mony. To test this speculation, we measured the effect of a number of possible sympathy factors across all the courts. In general, we found less apparent effect of sympathy factors than might have been expected. Factors such as the patient's age, medical condition, and gender did not consistently have an effect on the outcome. Indeed, we found the apparent anomaly that patients who are likely to die prevail significantly less often than those with some chance of death.

TABLE 8
OUTCOME BY LIKELIHOOD OF DEATH

	Highly Likely to Die	Some Risk of Death	Condition not Life Threatening
Percent of Patient Wins	51	79	51

p=0.004

We interpret this finding as suggesting that insurers are more cautious in denying coverage for life threatening conditions. If so, then cases where treatment is denied for life-threatening conditions would have much stronger support for the insurer than ordinary coverage denials. On the other hand, the fact that some risk of death creates a significant advantage for patients over no risk of death does suggest that some degree of patient sympathy is at work. Another measure of the seriousness of the patient's condition, which asked simply whether it was serious or mild, failed, however, to reveal a statistically significant association with the results in these cases.

TABLE 9
EFFECT ON OUTCOME OF THE SERIOUSNESS OF PATIENT'S CONDITION

	Serious	Mild or Not Ill
Percent of Patient Wins	61	49

p=0.140

A final area of relevant inquiry is the amount and quality of procedural protection that insurers give patients before a coverage denial becomes final. We did not extensively code for these procedural protections. However, we made a few noteworthy observations. We failed to identify any significant effect from providing patients a grievance procedure, arbitration, or a formal hearing,

even though there were forty-nine such cases found. Nevertheless, one indication that courts may be impressed with ample procedural protections is the high success rate in disputes arising from health insurance for federal government employees. In such cases, which follow an elaborate set of administrative procedures, courts found in favor of insurers much more often than under other types of public and private insurance.[13]

C. *Multiple-Factor Analysis*

Concerned that bivariate analysis could lead to misleading conclusions, we conducted multivariate analysis. We used logistic regression to develop a model of factors that predict whether a court will rule in favor of coverage. Multiple regression is a statistical technique that determines the effect of a single factor on a designated outcome, holding all other factors constant. Therefore, it is capable of separating out the commingled effects of many factors.

In performing a regression analysis, it is necessary to construct a model that contains only some of the many available factors. Including too many factors can reduce the utility of the analysis because several of the factors are likely to be measuring the same phenomenon at once. In such an event, most or all of the factors will appear insignificant, even though in combination they explain a great deal of the observed variation in outcome. In general, then, the objective in a regression model is to achieve "parsimony" by including the fewest factors that explain the most variation in the dependent variable with the greatest degree of statistical significance for each factor. As a result of this somewhat intuitive modeling exercise, we settled on the following set of factors that describe the legal forces at work in coverage disputes.

---

[13] *See* Table 3.

TABLE 10
PREDICTORS OF COURTS ORDERING COVERAGE

The effect of legal and other factors on the relative odds of patients winning a health insurance coverage dispute that involves a question of medical appropriateness. Asterisks indicate findings that meet standard tests for statistical significance, explained in note 10.

Variables	Odds Ratios	p value
Federal Appeals	.297	.005*
ERISA	.918	.828
Arbitrary & Capricious	1.51	.325
Insurer Discretion	.239	.001*
Specific Exclusion	.701	.499
Young (< 18 years)	.980	.968
Old (> 65 years)	2.42	.173
Female	1.78	.115
Some Chance of Death	3.03	.025*
Serious Condition	2.75	.011*
Treatment Worked	2.20	.060

The odds ratio column reports the relative odds of a patient winning when the factor is present, all other factors being equal. Thus, for instance, patients in federal appeals court have only thirty percent the odds of prevailing as those in other courts.

This regression analysis confirms many of the findings reported above. The most significant predictors of patient success are (1) not being in federal appeals court, (2) the contract not assigning discretion to the insurer, and (3) having some chance of death. Also, whether a patient suffers from a serious condition is a significant predictor of success in the multivariate analysis. Importantly, the effect of ERISA disappears in this multi-factor analysis, suggesting that its observed effect in the bivariate analysis is a consequence of other factors with which ERISA is associated, such as federal appeals court and insurer discretion. Other legal/contract factors such as the standard of review and the specificity of the exclusion language make no significant difference, nor do other patient factors such as age and gender. The final factor suggests, however, that courts are influenced by the unique attributes of the case before them. Patients for whom the treatment in question actually worked were twice as likely to prevail, and this difference comes close to meeting statistical significance for our sample size.

III. CONCLUSION

Courts appear to be receptive to patients' complaints that insurers have incorrectly denied them coverage. Patients win over half the time, even specific exclusions are frequently not enforced, and courts appear to be sympathetic to patients in a serious condition. Patients, however, prevail less frequently in federal appellate courts, and insurers gain a great advantage by writing their contracts to give them discretion over the meaning of general coverage language. The most troubling finding for patient protection in managed care settings is that very few disputes arising in this setting make their way into the courts. Either patients are not made aware of many of the coverage decisions that are implicitly being enforced, or they find it too expensive or too difficult to pursue their objections through the costly and time-consuming judicial process. Accordingly, calls for alternative dispute resolution mechanisms that are speedy and easy to access appear to be well founded.