New York State Documents

CALL No.: ***HEA 302-4 PATDL 90-33997***

TITLE: Patients, doctors, and lawyers: medical injury, malpractice litigation, and patient compensation in New York: a report..

AGENCY: Harvard Medical Practices Study.

CHECKLIST: October 1990

Original Document Scanned at:

■ 400 DPI ☐ Simplex ☐ Duplex

Original Document contained:

☐ Black & White Photos
☐ Colored Photos
☐ Colored Print (list color) _____
☐ Colored Paper (list color) _____
☐ Line Art, Graphs
☐ Oversized Pages -- reduced from _____ (original size)
☐ Text Only

Date Scanned: _____

This electronic document has been scanned by the New York State Library from a paper original and has been stored on optical media.

The New York State Library
Cultural Education Center
Albany, NY 12230

(MASTER.DOC. 9/99)

PATIENTS, DOCTORS, AND LAWYERS: MEDICAL INJURY, MALPRACTICE LITIGATION, AND PATIENT COMPENSATION IN NEW YORK

The Report of the
Harvard Medical Practice Study
To the State of New York

PATIENTS, DOCTORS, AND LAWYERS: MEDICAL INJURY, MALPRACTICE LITIGATION, AND PATIENT COMPENSATION IN NEW YORK

The Report of the

Harvard Medical Practice Study

To the State of New York

CONTENTS

TECHNICAL APPENDICES TO CHAPTERS

Chapter		
1-3)		none
4.)	4.II.1	Teaching Status of Hospitals in New York State
	4.II.2	Adjustment for Probability of Selecting Small Hospitals
5.)	5.III.1	Adjustment for Double-Counting of Adverse Events
	5.IV.1	Reliability and Validity of Physicians' Judgments and Validity of MRAs' Screens
	5.IV.2	Discoverability of Adverse Events through a Record Review
	5.V.1	Data Management, Verification, and Imputation
	5.V.2	Adjusting Sample Weights for Missing and Non-Reviewed Records
	5.V.3	Ratio Estimates and Variance Estimates for Implicitly Stratified Unequal Cluster Sampling Using SAS Proc SESUDAAN
	5.V.4	Classifying Patients into DRG Groups
	5.V.5	Standardized Rates and Computing in SAS RTI
	5.V.6	Comparing Follow-up and Reported Adverse Event Rates
	5.VI.1	Determining Disability Scores from Multiple Reviews
6.)		none
7.)	7.V.1	Matching Medical Record and Malpractice Claims
8.)	8.III.1	Survey Interview Process
	8.IV.1	Attributing Disability to Adverse Events
	8.IV.2	Adjusting Sample Weights for Unit Non-Response
	8.IV.3	Item Non-Response in the Patient Survey
	8.IV.4	Fringe Benefits
	8.IV.5	Revised Equivalence Scale for Urban Families of Different Size, Age, and Composition

TECHNICAL APPENDICES TO CHAPTERS

	8.IV.6	Tax Tables: Federal, State, City
	8.IV.7	Selection of the Control Group
	8.V.1	Weighted Average Number of Months of Equipment Use for Respondents Who Reported Needing Equipment: By Equipment Type and Respondent Category
9.)	9.III.1	Reasons for Specific Survey Questions
	9.III.2	Sampling for Mailed Survey
	9.III.3	Comparison of Responders and Non-Responders for the Mailed Survey
	9.III.4	Methods of Analysis of Survey Responses
	9.III.5	Comparisons by Specialty, Zone and Suit History for Factors Affecting Quality of Care. Influence of Various Disciplinary Actions, Changes in Practice
	9.IV.1	Sampling Interview Survey
	9.IV.2	Comparison of Respondents to Non-Respondents for the Interview Survey
10.)	10.III.1	Modifications to the Simple Example: Costs that Arise from Suit, Pricing in Medical Care, and the Customary Standard of Practice
	10.IV.1	Statistical and Econometric Issues
11.)		none

CONTENTS OF GENERAL APPENDIX
(separate volume)

Chapter

GLOSSARY

ADVERSE EVENT - an unintended injury caused by medical management rather than by the disease process. The injury is sufficiently serious to lead to prolongation of hospitalization or temporary or permanent impairment or disability in the patient. To be judged an AE, there must be a composite AE score of greater than 3.5. Close-call adverse events are cases with scores of 3.0 to 4.0. Low-threshold adverse events are cases with averaged scores greater than 1.0, up to and including 3.5.

CAUSATION - the attribution of a patient's disability to medical management rather than to the disease under treatment. The causation score reflects the reviewer's confidence in his/her judgment that medical management, rather than the disease process, caused the adverse event

CLAIM - a demand for tort compensation for injury and financial loss arising out of medical care

DISCOUNTING - the term used to describe the process of reducing future economic losses to their present value

IATROGENIC - any adverse condition in a patient resulting from treatment by a physician or surgeon

IBNR CLAIM - an "incurred but not reported" claim. The patient has suffered an injury for which he will eventually file a claim, but the provider has not recognized and reported the incident.

INDEX HOSPITALIZATION - a hospital stay sampled in the study

NEGLIGENCE - a failure on the part of the physician or other provider to provide reasonably careful treatment , i.e., treatment that normally should be expected from the practitioner usually caring for this kind of disease in the particular year in which the care was provided. In the Medical Practice Study, an average score of greater than 3.5

NO-FAULT - a system of compensation by health care providers for all injuries caused by medical management, irrespective of fault

NO-LIABILITY - a system whereby injured patients would pursue redress through the same public and private systems of loss insurance that are available to victims of any other disabling injury.

OBSERVATION or POTENTIAL CLAIM - a physician or hospital report to an insurer or agent that a bad outcome has occurred and might become the subject of litigation

POTENTIALLY COMPENSABLE EVENT - term used by the California Medical Insurance Feasibility Study to designate a disability caused by health care management

GLOSSARY

RELIABILITY - the reproducibility of a judgment. One measure comes from comparing the scores of multiple reviewers of the same medical record.

SELF-WEIGHTING DESIGN - each observation in the sample represents the same number of discharges. Raw rates and ratios calculated for the sample apply to the population.

SUIT - a civil complaint filed in court

TORT LAW - a system of legal liability for those who cause injuries due to substandard or negligent medical management

TWO-STAGE CLUSTER DESIGN - first, a random selection of hospitals (clusters) and second, a random selection of records within each chosen hospital

UNIVERSITY TEACHING HOSPITAL - 13 facilities in New York State designated by the state's medical schools as their primary clinical centers. Affiliate teaching hospitals are remaining hospitals with a minimum of 5 approved residency programs and 5 specialty hospitals with large numbers of residents on the staff.

VALIDITY - an estimate of the truth in a judgment. Measured by comparing judgments made with two or more methods. Construct validity is assessed by comparing one measurement process to another. The content validity of a process is evaluated by asking experts to examine it and to comment on its appropriateness.

ACRONYMS

ACOG - American College of Obstetricians and Gynecologists
AE - Adverse Event
AEAF - Adverse Event Analysis Form
AHA - American Hospital Association
AMA - American Medical Association
BLS - U.S. Bureau of Labor Statistics
CMA - California Medical Association
CPS - Current Population Survey
DDA - Discharge Data Abstract
DOH - Department of Health
DRG - Diagnostic (or Diagnosis) Related Group
HANYS - Hospital Association of New York State
JCAH - Joint Commission on Accreditation of Hospitals
MDC - Major Diagnostic Category
MLMIC - Medical Liability Mutual Insurance Company
MM - medical management
MMIA - Medical Malpractice Insurance Association
MPR - Mathematica Policy Research
MSA - Metropolitan Statistical Area
NAIC - National Association of Insurance Commissioners
NYHHC - New York City Health and Hospitals Corporation
OPMC - Office of Professional Medical Conduct
PCE - potentially compensable event
PRO - Peer Review Organization
PSU - primary sampling units
SPARCS - Statewide Planning and Research Cooperative System
SSDI - Social Security Disability Insurance
SU - sampling units
UBF - Uniform Billing Form

PREFACE

Concern about the medical malpractice problem and the tort litigation system as it deals with that problem led the then Deans Howard Hiatt of the Harvard School of Public Health and James Vorenberg of the Harvard Law School to bring together certain members of their faculties to form the Harvard Medical Practice Study in 1984. The complexity of the issues confronting legislative and executive bodies of government as well as the courts, physicians, lawyers, and society itself, and the paucity of facts that could illuminate those issues required the participation of members of both faculties and others if a comprehensive research program were to be carried out. An equally important requirement for such work was the sponsorship of a state government prepared to open to investigators hospital records, insurance records, and the participation of administrative units of hospitals, physicians, and several state and municipal agencies.

Benjamin Barnes and Harvey Fineberg of the School of Public Health and Paul Weiler of the Law School were members of the original study group. Weiler, who is also Chief Reporter of the American Law Institute's Tort Reform Project, has continued to serve as a principal architect and investigator. After Fineberg replaced Hiatt as Dean, he asked Hiatt, who is Professor of Medicine and whose background included nine years as Physician-in-Chief at a Harvard teaching hospital, to become a member of the group in 1985.

As the scope of the Study broadened, several colleagues from a range of disciplines joined it. William Hsiao, an economist at the School of Public Health, helped in the planning phase. Russell Localio, a lawyer-statistician, then Director of Research at the Risk Management Foundation, was recruited to manage the project and to work on medical record review design and execution

and claims data analysis. Ann Lawthers, a health policy analyst who was at Boston University, was initially administrative director and later coordinator and designer of the provider studies. Troyen Brennan, a lawyer-physician, member of the Division of General Medicine and Primary Care at Brigham & Women's Hospital, and a Lecturer at the Law School, became a senior member of the physician-reviewer group and a contributor to the provider studies. William G. Johnson, an economist at the Maxwell School of Syracuse University, assumed responsibility for the patient interview phase of the study. Nan Laird, a statistician at the School of Public Health, took charge of statistical design and methodology. Ken Thorpe, an economist at the School of Public Health, joined in the deterrence studies. Sol Fleishman and Howard Frazier, both internists, and Lucian Leape, and Lynn Peterson, both surgeons, were recruited to serve as senior physician reviewers for the record review portion of the study. In 1988, Leape, formerly chairman of the Department of Pediatric Surgery, Tufts Medical School, replaced Barnes as leader of the record review, and Joseph Newhouse, a health economist, formerly Director of the RAND-UCLA Center for Health Financing Studies and the new MacArthur Professor and head of Harvard's Division of Health Policy Research and Education, replaced Hsiao as leader of the econometric study. Liesi Hebert, an epidemiologist, joined the research team in 1989.

Consultants to the project included:

Floyd J. Fowler, Jr., Director of the Center for Survey Research, University of Massachusetts, who helped in planning the design of the hospital record survey.

Graham Kalton, Chairman of the Department of Biostatistics at the University of Michigan, who worked on the analysis of the survey sample.

Ruth Kilduff, Risk Manager at New England Medical Center, who helped design the survey on hospital injury prevention activities.

Donald Rubin, Head of the Department of Statistics at Harvard University, Alan Zaslavsky, Lecturer in Statistics, who assisted with the analysis of deterrence, and Theresa Dailey, who provided computational assistance for Chapter 10.

Members of the Harvard Medical Practice Study office who provided invaluable assistance during all phases of the study included: Sybil Carey, who provided administrative direction; Elaine Gebhardt and Steven Marcus, who assisted with computation and data management; Chris Braudaway-Bauman, Wendy Vander Hart, Sandra Lee Gould, and Robert Chaufornier, who provided secretarial assistance; and Roger Dempsey.

From the Metropolitan Studies Program, Maxwell School, Syracuse University, the following individuals assisted with the report: Bruce L. Riddle, academic computing specialist; Esther Gray and Martha Bonney, secretaries; Mary C. Daly, graduate research assistant; Linda McCarthy, research assistant; and Robert Guell, programmer.

A team from Mathematica Policy Research, Inc, of Princeton, New Jersey, under the leadership of Richard Strouse, carried out the patient interviews -- often under extremely difficult conditions -- very skillfully.

Support for the exploratory stages of the research came from the Klingenstein Fund of New York and a grant from the Risk Management Foundation of the Harvard Medical Institutions.

The relationship with the New York State Legislature and Department of Health under its Commissioner, Dr. David Axelrod, has been especially important. The Department's impartiality and commitment were crucial to that relationship, for the areas of medical malpractice and tort reform have been in urgent need of facts gathered and analyzed with methods that are scientifically sound. Also essential was the State's grant of complete confidentiality of information collected and the protection by New York law against subpoena of data.

Additions to, and further analyses of, our very large data base continue even as this report is submitted. As indicated in Chapters 7, 8, and 10 further material will later be available on litigation, compensation, and deterrence. That material will be included in the scholarly publications including the book that will emerge from this work.

Members of our group began with different views of the most promising ways to achieve reform. Some so regarded increased tort litigation, while others favored "no-fault" or other approaches. But as is necessary for every scientific enterprise, all agreed that our function was to gather the best possible empirical information. We emphasize this point for it has been suggested by some that the Study set out to prove that one approach was better than another. Rather, we believe we have succeeded in our goal--to gather unbiased information which will help inform and elevate the ongoing debate.

February 1990

EXECUTIVE SUMMARY

Introduction

The Harvard Medical Practice Study, carried out under contract to the State of New York, was designed to inform the policy debate now going on in New York and elsewhere about how society can best deal with its medical injuries and malpractice. To do so, we had to understand and isolate the key issues and assumptions that divide the protagonists of the current tort system, a reformed tort system, and no-fault alternatives. We have not prejudged the feasibility of any such no-fault program for injured patients, nor have we endorsed the criticisms that are made about present day malpractice litigation. Rather, we believe we have provided relevant empirical data that will permit informed judgments and sound policy-making concerning this complex area.

The Study had four principal components:

1. A population based measure of the incidence of injuries resulting from medical interventions, which we called "adverse events," and a determination of the percentage of such events that resulted from fault or negligence of the physician or other provider.

2. A determination of the percentage of adverse events, both negligent and non-negligent, that led to claims and suits. In addition, we obtained information about the numbers of claims and suits by patients in whose hospital records we found no evidence of injury.

3. Measures of the costs of medical expenses, lost wages, and lost household production to the victims of medical injuries and to their families, and their compensation for such losses under current arrangements.

4. Estimates of the degree to which variations in the threat of litigation affected the incidence of adverse events.

The following summarizes some of our methods and major findings.

1. The incidence of adverse events

The hospital medical record review was key to estimating the incidence of adverse events associated with medical management. The record review focused on two critical issues: causation and negligence. We asked, "Was the patient's condition attributable to medical management rather than to the disease under treatment (causation)? Was negligence involved?"

In addition to establishing causation and negligence, we determined where injuries occurred, the types of injury and then the magnitude of disability experienced.

The review was conducted by teams of trained medical record administrators and nurses for the screening phase, and board-certified physicians for the physician-review phase.

Methods were devised to resolve the logistic problems that arose because of the infrequency of adverse events: we found efficient and reliable ways to sift through thousands of medical records to find the few that indicated the patient disability caused by medical management. We also developed ways to deal with the methodologic problems that arose: the medical record administrators had to make valid judgments regarding the presence of screening criteria and physicians had to make valid and reliable judgments about whether a patient's injury resulted at least in part from medical management, and, if so, whether management failed to meet a standard of medical care.

In order to make our results generalizable to the entire population of hospital discharges in New York, we drew a probability sample of more than 31,000 hospital records. Our ability to obtain such a sample was made possible by the availability of the Statewide Planning and Research Cooperative

System (SPARCS) data system. The basic sampling design of the Study was an implicitly stratified, systematic, two-stage cluster sample of discharges. We first selected hospitals with probabilities proportional to the number of non-psychiatric discharges and then secured the cooperation of all 51 hospitals selected. Records within hospitals were selected with three different sampling frequencies determined by patient age and diagnosis-related group (DRG). Using SPARCS information on patient discharges, we drew a sample with a distribution that conformed closely to the population on important hospital and patient characteristics.

We analyzed 30,121 (96%) of the 31,429 records selected for the study sample. After preliminary screening, physicians reviewed 7,743 records, from which a total of 1,133 adverse events were identified that occurred as a result of medical management in the hospital or required hospitalization for treatment. Of this group, 280 were judged to result from negligent care. Weighting these figures according to the sample plan, we estimated the incidence of adverse events for hospitalizations in New York in 1984 to be 3.7%, or a total of 98,609. Of these, 27.6%, 27,179 cases, or 1.0% of all hospital discharges, were due to negligence.

Physician confidence in the judgments of causation of adverse events spanned a broad range, but only 1.3% of all discharges were in the close-call range (defined as a confidence in causation of just under or just over 50-50). An even smaller fraction, 0.7% of discharges were close-call negligent adverse events, but they constituted a larger proportion of total negligent adverse events.

The majority of adverse events (57%) resulted in minimal and transient disability, but 14% of patients died at least in part as a result of their adverse event, and in another 9% the resultant disability lasted longer than 6 months. Based on these figures, we estimated that about 2,500 cases of permanent total

disability resulted from medical injury in New York hospitals in 1984. Further, we found evidence that medical injury contributed at least in part to the deaths of more than 13,000 patients in that year. Many of the deaths occurred in patients who had greatly shortened life expectancies from their underlying diseases, however. Negligent adverse events resulted, overall, in greater disability than did non-negligent events and were associated with 51% of all deaths from medical injury.

Risk factors

The risk of sustaining an adverse event increased with age. When rates were standardized for DRG level, persons over 65 years had twice the chance of sustaining an adverse event of those in the 16-44 years group. Newborns had half the adverse event rate of the 16-44 years group. The percent of adverse events resulting from negligence was increased in elderly patients. We found no gender differences in adverse event or negligence rates. Although the rates were higher in the self-pay group than in the insured categories, the differences were not significant. Blacks had higher rates of adverse events and adverse events resulting from negligence, but these differences overall were not significant. However, higher rates of adverse events and negligent events were found in hospitals that served a higher proportion of minority patients. At hospitals that cared for a mix of white and minority patients, blacks and whites had nearly identical rates.

Adverse event rates varied 10-fold between individual hospitals, when standardized for age and DRG level. Although standardized adverse event and negligence rates for small hospitals (fewer than 8,000 discharges/year) were less than for larger hospitals, these differences were not significant. Hospital ownership (private, non-profit, or government) also was not associated with significantly different rates of adverse events. The fraction of adverse events due to negligence in

government hospitals was 50% higher than in non-profit institutions, however, and three times that in proprietary hospitals. These differences were significant. The standardized rate of adverse events in upstate, non-MSA hospitals was one-third that of upstate metropolitan hospitals and less than one-fourth that in New York City. These differences were highly significant. The percent of adverse events due to negligence was not significantly different across regions. Non-teaching hospitals had half the adverse event rates of university or affiliated teaching hospitals, but university teaching hospitals had rates of negligence that were less than half those of the non-teaching or affiliated hospitals.

The nature of adverse events

Nearly half (47%) of all adverse events occurred in patients undergoing surgery, but the percent caused by negligence was lower than for non-surgical adverse events (17% vs 37%). Adverse events resulting from errors in diagnosis and in non-invasive treatment were judged to be due to negligence in over three-fourths of patients. Falls were considered due to negligence in 45% of instances.

The high rate of adverse events in patients over 65 years occurred in three categories: non-technical postoperative complications, complications of non-invasive therapy, and falls. A larger proportion of adverse events in younger patients was due to surgical failures. The operating room was the site of management for the highest fraction of adverse events, but relatively few of these were negligent. On the other hand, most (70%) adverse events in the emergency room resulted from negligence.

The most common type of error resulting in an adverse event was that involved in performing a procedure, but diagnostic errors and prevention errors were more likely to be judged negligent, and to result in serious disability.

The more severe the degree of negligence the greater the likelihood of resultant serious disability (moderate impairment with recovery taking more than six months, permanent disability, or death).

2. Litigation data

We estimated that the incidence of malpractice claims filed by patients for the study year was between 2,967 and 3,888. Using these figures, together with the projected statewide number of injuries from medical negligence during the same period, we estimated that eight times as many patients suffered an injury from negligence as filed a malpractice claim in New York State. About 16 times as many patients suffered an injury from negligence as received compensation from the tort liability system.

These aggregate estimates understate the true size of the gap between the frequency of malpractice claims and the incidence of adverse events caused by negligence. When we identified the malpractice claims actually filed by patients in our sample and reviewed the judgments of our physician reviewers, we found that many cases in litigation were brought by patients in whose records we found no evidence of negligence or even of adverse events. Because the legal system has not yet resolved many of these cases, we do not have the information that would permit an assessment of the success of the tort litigation system in screening out claims with no negligence.

Confining our analysis to the adverse events that involved strong or certain evidence of negligence, however, we estimate that 12,859 injuries from medical negligence did not lead to malpractice claims. Of these injuries, 22% (2,833) occurred in patients under age 70 years who suffered moderate or greater incapacity. Our projections suggest that if this group of

patients had litigated, the malpractice claims frequency for year 1984 would have increased by 75%.

3. Economic Consequences of Medical Injury

Having documented from the medical records survey which patients were injured, and from the litigation survey which patients filed tort suits, we used the patient survey to determine from the patients themselves what losses they suffered as a result of these injuries and what compensation they received from non-tort sources. For that purpose we divided our patient sample into five categories -- worker, homemaker, child, retired, and disabled -- and assembled data about lost wages and fringe benefits, medical costs, lost household production, and levels of physical and functional impairment. Our data for that final category have not been analyzed for this Report.

We faced two major difficulties in this survey. First, we had to locate, in 1989, people who had been hospitalized in 1984 in order to interview them about their experience since 1984. In fact, we were successful in finding and interviewing 71% of all injured patients, a response rate which is quite respectable for a survey of this type.

Our second problem was how to disentangle the effects of the adverse event itself from those that were properly attributable to the underlying illness, which itself would naturally be expected to entail considerable medical costs, time off work, and inability to perform normal household tasks. Two different strategies were devised for this purpose. One was to interview a control group of uninjured patients who were matched with our "experimental" group on the relevant dimensions, thus permitting econometric analysis of the precise difference which the iatrogenic injury made in the aggregate economic experience of the two groups. While we have collected all the data for the two groups, we have not completed this analysis for purpose of presentation in this Report.

Instead our primary focus has been on an alternative method -- estimating the compensable losses that might be paid under a hypothetical no-fault plan in which each patient's experience was assessed individually (as would have to be done in a real no-fault program), and then totaled. For that purpose we had to make a number of assumptions about program design: two important ones are noted here. First, all financial losses and compensation received during the first six months from hospital admission were deleted. These short-term losses are likely reimbursed from other sources (e.g., sick pay for time off work). Further, this reduces the number of cases in which disentangling the effect of the injury from the underlying illness may be very difficult. Second, we assumed that a no-fault patient compensation scheme would involve a second insurer, standing behind primary sources of general medical or disability insurance.

Our key findings with respect to these two criteria were that the bulk of disabilities were of short duration -- e.g., 42% of absences from work lasted for less than a month and 76% lasted less than six months. However, the average economic losses were much larger in the smaller number of more serious or fatal disabilities. With respect to these longer-lasting disabilities, more than 85% of the medical bills were covered by some form of health insurance, but only 20% of the lost earnings, and no detectable portion of lost household production.

Our ultimate finding is that the present discounted value of the net compensable losses (past and future) suffered by patients injured in New York hospitals in 1984 amounted to $894 million (in 1989 dollars). These compensable losses consisted of $285 million in lost wages and fringe benefits, $103 million in uninsured medical costs, and $506 million in lost household production (the latter having been valued at the market wages earned by the working women in our patient cohort).

To provide some perspective for these figures, the malpractice premiums paid by New York doctors and hospitals in 1988 amounted to $850 million. When one includes the amount spent by self-insured hospitals and the health care organizations, the total malpractice insurance burden is over $1 billion. However, these tort costs incorporate two major factors not reflected in our estimate. One is damage for pain and suffering, which typically are not compensated under no-fault programs. The other component is administrative and legal expenses which definitely would be a significant factor over and above the patient's economic losses. The administrative share of claims costs in no-fault workers compensation is usually estimated to be around 20%, though we believe it would be somewhat higher for no-fault patient compensation.

Since the sample of injured and interviewed patients in our different categories was rather small despite the relatively large sample of 31,000 hospitalizations, the confidence intervals surrounding our point estimates are large: the figures might be as much as 50% less or 100% more than those presented.. On the other hand, the estimate of net wage losses and medical costs -- these being the items typically covered by a no-fault scheme, and even then not in full fault - totalled just $335 million. Thus, there is considerable room within the current tort "envelope" to adjust even for an outcome at the highly improbable outer limit of these confidence estimates.

4. Malpractice Litigation and Deterrence

We examined the presumed deterrent effects of the tort system in two ways -- a series of physician surveys as well as an econometric study that compared the rates of adverse events and negligent adverse events, on the one hand, with the threat of a claim on the other.

The physician surveys revealed that the overall perceived risk of being sued in a given year was 20%, approximately 3 times the actual risk of being sued. The perceived risk of suit for

negligent care was about 60%, a figure substantially greater than the actual risk of litigation from injuries caused by negligence. Additionally, perceived risk was significantly greater for high-risk specialties such as obstetrics, orthopedics and neurosurgery and for physicians in Nassau and Suffolk counties, lending credence to the responses.

Physicians who perceived themselves to be at greater risk of suit said that in the past ten years they had ordered more tests and procedures and reduced their practice scope more than had their colleagues with perceived risk.

The tort system's deterrence signal to physicians appeared mixed. For example, physicians often considered the severity of punishment to depend on whether a case went to trial or whether the media publicized it. The evidence was not clear, however, on whether the severity of the punishment and the actual transgression were related: most physicians perceived their suits as having arisen from circumstances beyond their control. Many seemed to believe that the threat of the tort system was too broad and lacked specificity.

Although physicians believed they practiced medicine defensively, they did not report long-term changes in their practice patterns as the result of a specific suit. Thus, it was not clear whether defensive medicine resulted from the malpractice environment or from other factors such as advances in the science and technology of medicine, changes in societal expectations as to what constitutes an appropriate level of care, or changes in Peer Review Organization (PRO), state and hospital requirements, or a combination of factors.

Another important finding concerned physician attitudes about iatrogenic injury and negligence. Physicians tended to equate a finding of negligence with a judgment of incompetence. Thus, although willing to admit that all doctors make mistakes, physicians were often unwilling to label substandard care as negligent and were opposed to compensation for iatrogenic injury.

The final part of our study examined the relationship between variations in claims rates and variationsin cost and in injury rates in the sample of study hospitals. We found some evidence that total cost per discharge was greater in hospitals that faced higher claims rates, although the relationship that we estimated was sensitive to how we specified the relationship. Even conceding that there is an effect on cost, however, does not tell us whether the effect is good or bad. On the one hand, greater efforts to prevent injuries or ameliorate the consequences of those that occur may well require greater resources. On the other hand, additional resources in response to a greater threat may simply represent wasteful defensive medicine and not contribute to a reduction in patient injuries.

The important test, therefore, is whether hospitals that face higher claims rates actually do exhibit lower injury rates. We find no evidence that they do, but the precision of our estimates is not good, and we cannot rule out the possibility that there are in fact substantially reduced rates of injuries at the hospitals in our sample with higher claims rates. More specifically, the point estimate relating injuries to claims is actually positive in most specifications and never close to significantly negative. However, the confidence intervals around the coefficient include values that would demonstrate substantial deterrence.

We illustrate how our data cannot rule out a substantial deterrent effect by choosing one of the relationships we estimated, that for the probability that an adverse event is negligent, controlling for a number of other hospital characteristics. The point estimate of the claims variable is slightly positive; however, if we reduce the point estimate by approximately one standard error, it shows substantial deterrence. In quantitative terms, the reduced estimate would suggest that, other things equal, hospitals in the highest quartile of claims rates would have about 24% fewer negligent

events (conditional upon an adverse event) as those in the lowest quartile.

Moreover, there may be a bias in our results toward showing no deterrent effect. Our goal was to determine whether there is a negative relationship between claims rates and injuries, but hospitals and physicians that have higher injury rates may have more claims filed against them. This possible positive relationship between injuries and claims would tend to mask any true deterrent effect. We have tested for this bias and do not find any evidence of it, but our test could simply be failing to detect it.

Finally, even if we had been able to conclude that our data ruled out all but a negligible deterrent effect, we could not conclude that abolishing the tort system would have no effect on injury rates. All the hospitals in our sample faced some threat of a claim if an injury occurred, and the most we could hope to learn was the effect on injury rates of variation in that threat. Abolishing the tort system could reduce that threat to zero (depending on what, if anything replaced it), and we cannot learn from our data what the effect of that might be.

Chapter 1

THE MALPRACTICE SETTING WITHIN NEW YORK STATE

The Harvard Medical Practice Study was commissioned by the Department of Health of New York State as an offshoot of a major legislative reform in 1986 of the state's Medical and Dental Malpractice and Professional Conduct Act. Section 39 of that Act required a study of the "Extent of medical and dental injury regardless of fault, the extent to which such injuries are compensated by sources of compensation outside the liability system, and an analysis of alternative insurance or liability mechanisms ... including the feasibility of medical compensation systems that would compensate victims irrespective of fault." Both the 1986 legislation and this Study are an outgrowth of a larger history of the medical malpractice system within New York. Thus it is useful to sketch briefly some of that background as a prelude to the detailed presentation of the nature, execution and results of the Harvard Medical Practice Study.

As we complete this report in the fall of 1989, there is nowhere near the sense of immediate crisis enveloping malpractice insurance in the state as there was when we were first contacted by the Commissioner in the summer of 1986. For the two or three years before and after that date, the state's malpractice carriers regularly sought steep premium increases and were usually granted substantial price hikes. This year, rates across the state will generally not change (although physicians in the high litigation risk counties in and around New York City will still face a slight increase).[1] In that respect, New York is in

---

[1] Medical Liability Mutual Insurance Company (MLMIC), the largest insurer in the state, and the Medical Malpractice Insurance Association (MMIA), the joint underwriting pool, requested and received no overall premium increases. Likewise, Frontier, which covers 9% of the physicians, received no increase. Physicians' Reciprocal and Group Council, which collectively insures 31% of the physicians in the state, received a 5% overall increase and a 4% surcharge, to compensate for rate deficiencies of years past (personal communications from Mr. Henry Lauer, Insurance Department, July 6, 1989). MLMIC requested a 5% rate increase for some of its insured physicians and received a 1.6% raise for them. A.M. Kaufman & R.S. Biondi, Medical Liability Mutual Insurance Company: 1989-90 Rate Analysis (May 1, 1989) (on file at the Property and Casualty Bureau, New York State Insurance Dept., 160 West Broadway, New York, NY 10013).

accord with states across the nation whose malpractice premiums have stabilized and even subsided somewhat.[2]

Even if we have reached a reasonably level plateau for the moment, however, its actual height looks rather astonishing, at least to a casual observer. The current cost of malpractice insurance protection for up to $3 million per claim for a Long Island neurosurgeon, an orthopedic surgeon or an obstetrician ranges from $160,000 to $200,000/year, and in New York City and its adjacent counties the costs are not far behind.[3] The average premiums paid by or on behalf of the 25,000 insured practitioners in the state for just the basic $1 million coverage per claim is roughly $23,000/year. The two excess layers of $1 million coverage each, which provide total coverage of $3 million per claim, add 80% to this basic premium charge. And at least until recently, there was reason to believe that these premium levels were considerably lower than a purely actuarial judgment would suggest because premium increases had been restrained by the Superintendent of Insurance acting under the state legislation which itself has been the target of a number of court challenges in the last two or three years.[4]

At the moment, though, the medical liability insurance system has reached a point of stability. As we noted above, the rate increases established by the Superintendent of Insurance for

---

[2] Schiffman, Medical Malpractice Insurance Rates Fall: Drop in Number of Claims Cuts Insurers' Costs, Wall St. J, Apr. 28, 1989 B1; Freudenheim, Costs of Medical Malpractice Drop After 11-Year Climb, N.Y. Times, June 11, 1989.

[3]. These figures are drawn from the rate schedules issued for July 1, 1988 - June 30, 1989 by MLMIC and MMIA.

[4] Legislative activity at the end of the 1988 session makes clear that the state lawmakers hope that past deficiencies in premiums will be recouped by a combination of prospective surcharges and delayed effect of tort reform legislation.

> The legislature further finds that it is appropriate to clarify the nature of the authority of the superintendent of insurance to establish rates in view of a judicial decision that has misconstrued the intent of the legislature so that the superintendent of insurance will have, with certainty, the full authority contemplated by the legislature to establish an adequate rate taking into account the ability to impose a surcharge as well as such superintendent's judgment as to the prospective effects of legislative and regulatory reform. 1988 N.Y. Laws ch. 184, §1.

the policy year beginning July 1, 1989 keep premium levels essentially unchanged, in stark contrast to the experience of the previous half dozen years.

At their current level, these New York malpractice insurance rates are not the highest in the nation: Florida has that rather dubious distinction. But with the exception of Florida (and Michigan), the cost of malpractice coverage in New York is much higher than it is in other states where such coverage also is provided by carriers owned by the doctors themselves, including such adjacent states as New Jersey and Connecticut.[5] Indeed, even within New York itself, this situation is of recent vintage. In 1949, the average premium paid by New York doctors was just $65 a year. It was still only about $250 in 1965, had jumped to $3,000 by 1975, and now is about $23,000 (and recall that the latter figure is just for the basic coverage, with the more sensible protection of up to $3 million per claim costing nearly $40,000 on average).[6]

How have events unfolded to bring us to this point? Table 1.1 graphically depicts the underlying trends which, from the point of view of the Superintendent of Insurance, have been driving malpractice premium levels within the state.[7]

---

5. Rate comparisons for Florida, Michigan, New York, Alaska, Arizona, Georgia, Hawaii, and California are tabulated and graphed in Cohn, Tort Reform: Past, Present, and Future; 74 Am. Coll. Surgeons Bull. 13-17 (1989).

6. These historical comparisons within New York State are drawn from the Report of the New York State Insurance Department on Medical Malpractice, A Balanced Prescription for Change (herein after A Balanced Prescription for Change) 51 (1988) (for the 1949 and the 1987 figures) and from the Report of the State of New York's Special Advisory Panel on Medical Malpractice 243 (1976) (for the 1965 and 1975 figures).

7. The figures in Table 1.1 and the analyses in the text are drawn from, note 6 supra, A Balanced Prescription for Change 47-64 (1988). In Chapter 7 of our Report, we use a different definition of malpractice claims. While this provides a somewhat different picture of the actual claims ratio, it does not materially alter the Insurance Department's presentation of the historical trend.

TABLE 1.1

EVOLUTION OF MALPRACTICE CLAIMS AND PREMIUMS IN NEW YORK STATE

Category	1955	1965	1974	1980	1985	% Increase 1955-1985
Frequency of Claims(1)	1.9	2.6	7.2	8.0	8.4	340%
Limited Severity(2)	$6,400	$28,000	$140,000	$290,000	$480,000	7400%
Unlimited Severity(3)			(1976) $205,000	$390,000	$1,050,000	xxx
Pure Premium Per Doctor(4)	$123	$741	$10,000	$24,000	$38,000	31,000%

* All figures (rounded off) derived from Insurance Department Report on Medical Malpractice.

(1). Number of paid claims per 100 insured doctors per year
(2). Average payments on successful claims from primary coverage (up to $1 million per claim)
(3). Average payments on successful claims from both primary and excess coverage
(4). Total annual payments to patient-claimants divided by the number of insured doctors in that year

The following are the key points to be discerned from this Table.

(i) Whereas in 1955 roughly two claims were paid per 100 doctors, by 1985 nearly eight and one-half claims were paid. Because fewer than half the claims made by patients lead to payments by doctors or their insurers, New York doctors now experience one claim for every sixteen doctors every year -- or, over a period of six years, an average of one claim for each doctor.

(ii) The amount of money paid on average for each successful claim under the doctor's basic policy coverage amounted to just over $6,000 for claims arising out of treatment in 1955, and this amount is

projected to be nearly $500,000 per claim arising from 1985 treatment.

(iii) That last figure is actually artificially low because it reflects only payments made under the primary policy whose $1 million ceiling has remained stable in nominal terms for over a decade, while the number of court judgments and out-of-court settlements over $1 million has spiraled upward. That is why more and more doctors now find it necessary to have excess coverage of an additional $1 million or $2 million per claim in order to protect themselves against occasional catastrophic injury to their patients. Whereas in 1976 only one quarter of the total payments came from the excess insurance coverage, the Superintendent of Insurance projects that two thirds of total payments will come from excess coverage for claims based on treatment in 1987. When one counts payments made for these policies, the Superintendent's projected average payout for 1985 incidents is actually over $1 million per successful claim -- versus the $6,000 a claim just three decades earlier.

(iv) When one combines these trends in average frequency and severity, the amount paid by malpractice insurers (actually just by the primary insurer) to injured patients (without counting the insurer's expenses in adjusting these claims), divided by the number of insured doctors in the state (thus controlling for changes in the doctor population), rose from just $123 in 1955 to over $38,000 in 1985. And that 300-fold increase in just 30 years would be far higher if one also counted the money spent by the excess insurers on claims.

At the same time we must make clear that the figures from

later years incorporated in Table 1.1 do include mostly projected rather than actual claims payments. The reason why projected losses must be included in the total losses incurred by the insurer is that considerable time elapses between the time at which a patient is treated, a disabling injury eventually manifests itself, a lawyer is consulted and a claim is filed, and the claim is ultimately resolved either by settlement or trial verdict. Medical malpractice insurance has a particularly "long tail" in its ultimate disposition of tort claims stemming from a particular treatment year: on an average of 6 years, and for the more serious cases often 10 years or longer. Thus, the insurer's actuaries must extrapolate into the future the trends and patterns that the firms are currently experiencing in order to determine what premiums should be charged and what reserves should be set aside.

It is reasonable to ask, then, whether the latest figures reported in Table 1.1, which, especially for 1985, are based on claims and settlements that might not be paid until well into the 1990s, have been artificially inflated by alarmist judgments by the actuaries. One will not be able to make a definitive judgment on that score until we reach the end of the tail from the 1985 treatment year. However, the sudden flattening in this year's premiums does indicate that insurers have markedly moderated their projections of future trends in claims frequency and severity.

Still, experience suggests that we should not be unduly optimistic about the future. The report of the Superintendent of Insurance[8] contains a well-established test for that hypothesis that we can apply to prior years that are far enough back in time to give us a reasonably complete picture of the claims actually experienced from that year. For accident treatment years beginning with 1959, the amount of money received

---

[8]. A Balanced Prescription for Change, note 6 supra, 68-74

by insurers in premiums and in investment earnings from reserves is accounted for separately, and then so-called "withdrawals" are made from that account for all expenditures on awards and adjustment costs for claims for that treatment or accident year. In every accident year from 1959 through 1979, the amount of funds withdrawn has exceeded the total amount deposited at a point in time when a considerable number of sizable claims had not yet been settled and payments made. The costs of the claims that remained from those insured years in which the actuaries were too optimistic had to be "borrowed" from premiums paid by doctors in subsequent years. The sobering lesson from this exercise is that the actual experience of claims in the 1990s, based on medical treatment in the 1980s, may not turn out any better than did claims made in the 1980s, based on treatment in the 1970s. (As shown in Chapter 7, the findings of this study will offer a more informed basis for judging whether and how much room for growth there is in malpractice litigation and premiums.)

In any event, there is no doubt that larger and larger sums of money are being expended within the state's malpractice liability insurance system. We should be under no illusion, however, that inordinate sums are being won by the typical injured patient.[9]

We noted earlier that an average of six years is required for any payment to be made following the injury (3.5 years from the time the claim was filed). Fewer than half of all such claims produce any payment at all (but fortunately both the success rate and average payment are significantly higher for those patients who suffer the more severely disabling injuries). Depending on the size of the payout, up to one third of the amounts received by the patient must be paid to the latter's lawyer, who takes these cases on a contingent basis, and thus is

[9]. These assertions about operation of the medical liability system in New York State are based on a detailed analysis by the Department of Health of all malpractice claims closed from 1980 through 1983, a study reported in Appendix B of N.Y. Department of Health, Monitoring Health Care Quality (1988)

paid nothing in more than half of the cases. And as shown later in this Report, even more patients who are actually injured by medical treatment never bring a potentially valid claim into the tort system, notwithstanding the spiraling frequency and severity rates we have recounted.

Nor are these injuries and claims a product of just a handful of especially accident-prone doctors. The Department of Health's review of all reported[10] claims closed against physicians from 1980 through 1983 found that only 131 doctors (just 0.36% of the 35,552 physicians practicing during that period) had been the subject of more than two paid tort claims during those four years. As well, the vast bulk of these multiple-claims doctors practiced in the higher-risk specialties of surgery and obstetrics. If one excluded a single osteopath who paid 239 claims for negligent hair transplants, claims against this particular group of physicians amounted to less than 10% of the total paid claims against all New York doctors in that period. Clearly, modern medicine is a risky enterprise. Momentary, inadvertent mistakes can be made by even the most careful and concerned doctors, just as by their counterparts in any other walk of life. But when a doctor makes such a mistake, the unfortunate result is that a patient may be hurt, often severely, and the tort suits that result ultimately must be met by the malpractice insurance system whose costs we have seen soaring upwards.

The trends here have been quite visible to the state government, which has responded with significant legislative reforms of the system, initially in 1975-1976 amid the state's first malpractice crisis and again in 1985-1986 when the problem re-emerged at an even starker level. Additional legislation on

---

[10] The extent of reporting as required by statute and regulation has been the subject of a great deal of discussion among the Medical Practice Study, the Department of Health, and the Insurance Department. See Interim Report, 39-40, 49. For the efforts that the Study has undertaken to achieve complete reporting for this report, see Chapter 7.

professional liability and insurance has been passed in almost every legislative session from 1975 to 1989.

These statutory changes have addressed each of the three distinct components of the overall malpractice system. One is the institution of liability insurance, which has been producing the ever increasing premium charges to doctors and hospitals in order to accumulate the pool of funds used to pay injured victims. The second component is the tort litigation process, which determines which injured patients will be able to lodge successful claims and how much compensation they will receive. The third is the health care system itself, in which patients are treated by doctors or other providers: sometimes this treatment aggravates rather than ameliorates the patient's condition, thus leading some of these patients to have recourse to the tort system to secure monetary relief for their injuries. Each of these components of the malpractice setting has received sustained attention from the state government over the last 15 years.

Without going into detail, it is useful to summarize the range of measures that have already been tried within New York. As regards the insurance component of the system, the Legislature responded in 1975 to a serious threat caused by unavailability of insurance for doctors by forcing the creation of the Medical Malpractice Insurance Association (MMIA), a joint underwriting association comprising all liability insurers doing business within the state.[11] The government also licensed the new Medical Liability Mutual Insurance Company (MLMIC), a doctor-controlled insurer that now provides the bulk of primary coverage to New York physicians. Then in 1986, in order to make such insurance more affordable, the Legislature required insurers to write only so-called claims-made policies to physicians who were under the

[11] 1975 N.Y. Laws ch. 109, sec 17 (codified at N.Y. Insurance Law sec 5501 et seq. (Consol. 1985), as amended (cum. supp. 1988).

age of 59 and who did not already have an occurrence policy in effect.[12] By limiting liability to claims made during the year the policy was in effect, the law would achieve a onetime reduction in overall insurance rates, although eventually the premiums required for claims-made and occurrence policies would be roughly the same. Legislation also required hospitals to purchase a first layer of excess insurance for all their attending physicians,[13] thus shifting part of the cost from physicians to hospitals and thence to health insurers through the reimbursement mechanisms.

Insurance regulations issued in 1983 permitted certain medical institutions to "channel" all the insurance coverage of their physicians into a single hospital policy (with the doctors paying for their share of the somewhat cheaper single coverage).[14] The Legislature has mandated development of a merit rating program that would charge higher amounts to doctors who experienced abnormally high claims and payments.[15] More recently, the Legislature and the Superintendent of Insurance have placed stringent limits upon the size of annual premium increases, while providing for surcharges in future years in order to make up any resulting deficits.[16]

Reform of the tort litigation system has been directed at each of its three key points: access of claimants; determination of liability; assessment of damages. Measures to contain the

---

[12] 1985 N.Y. Laws ch. 266, sec. 9

[13] 1985 N.Y. Laws ch. 294, sec. 19 (McKinney 1985), as amended, 1986 N.Y. Laws ch. 266 sec. 18 (McKinney 1986), 1988 N.Y. Laws ch. 184, sec. 18 (McKinney 1988)

[14] N.Y. Comp. Codes R. & Regs. tit. 11, sec 16 (1983)

[15] N.Y. Ins. Law sec. 2343(d) (Consol. cum supp 1988)

[16] As we noted earlier, this latest statutory policy produced legal challenges by the insurance industry, which eventually were rejected by the state courts (Medical Malpractice Insurance Association v. Supt. of Ins., 72 N.Y.2d 753, 537 N.Y.S.2d 1 (1988)) but are still active in the federal courts (Alliance of American Insurers v. Cuomo, 854 F.2d 591 (2d Cir. 1988)). By agreement of the parties, the federal action has been stayed for one year. Stipulation and Order. Alliance of American Insurers v. Cuomo, 87 Civ. 0169, June 27, 1989.

entry of claims into this process include somewhat shorter statutes of limitations, a so-called good samaritan law for those who provide medical treatment in an emergency outside of a hospital or office, mandatory screening panels, voluntary arbitration, restrictions on contingent fees chargeable by the patient's attorney, and monetary penalties against lawyers who file spurious claims.

Once a case is within the tort system, the common law principles for appraising the merits of the injured patient's claims -- in particular, the reliance on customary medical practice to define the standard of care -- are actually somewhat more favorable to the doctor than they are to tort defendants in other injury contexts. These doctrines have not been touched by the Legislature (with the exception of codification of the doctrine of informed consent and its exceptions).

But the standards for measuring and awarding damages to successful plaintiffs have been modified in a variety of ways: by requiring itemization of each ingredient in the jury's award; by offsetting this award with past and future payments from collateral sources of compensation for that injury; by jury consideration of the non-taxability of the damage verdict; by periodic payments of large monetary awards, and by terminating any such payments for projected medical costs and pain and suffering if the victim dies prematurely.

Of the entire array of standard tort reform proposals, the only one not adopted in New York is a statutory cap on the damage award, in particular for nonpecuniary losses. While some evidence indicates that such caps do contain some of the increase in tort suits and insurance premiums, they do so only at considerable injustice to the most seriously injured victims whose claims are the only ones that are constrained by a fixed monetary ceiling. In lieu of that broad brush approach, New York has enacted more specific measures to minimize the burden on defendants. First, the law now precludes the use of "joint and

several liability" to collect damages for pain and suffering from the "deep pocket" of the well-insured defendant when that defendant is less than 50% liable for the damages.[17] Second, the Legislature invited the appellate courts to review and revise any jury awards that "deviate materially from what would be reasonable compensation" (as contrasted with the much more limited appellate scrutiny under the previous judicial criterion of whether the jury award "shocked the conscience of the court").

Actually, a well-known fact among lawyers engaged in malpractice litigation in New York is that even without a statutory cap on damages and prior to the effective date of the statute requiring closer appellate scrutiny of damage awards, the appellate courts regularly and often substantially reduced jury verdicts in malpractice cases. For several years the courts did so with little explanation in their opinions but with apparent serious concern for the tendency of some jury awards to exceed the levels of damages supported by the evidence or to transgress general notions of fairness. Although we have no reports of a systematic, complete analysis of the reductions, appellate courts in the cases footnoted here reduced verdicts 50 to 75% from what the jury had awarded.[18] Also known, however, is the constant upward trend in the largest verdicts that the appellate division has permitted. Each new increase in the largest judgment creates pressure on the insurers to increase settlement offers in cases

---

[17] N.Y. Civ. Prac L. & R. sec. 1601 (Consol. cum. supp. 1988)

[18] Sakin v. Fryman, 538 N.Y.S.2d 35, App. Div. (2d Dept. 1989) (reduction from $255,000 to $105,000); Gines v. Maimonides Medical Ctr., 137 App. Div.2d 582, 524 N.Y.S.2d 473 (1988)(damages for injury to sciatic nerve reduced from $1.5m to $350,000); Marmo v. Southside Hospital, 533 N.Y.S.2d 402 App. (Div. 1988) (where defendant conceded liability and 33-year-old patient sustained brain damage, court reduced pain and suffering from $3m to $1m and compensatory damages from $1.1m to $550,000); Beverly H. v. Jewish Hospital and Medical Cntr., 135 App. Div.2d 497, 521 N.Y.S.2d 238 (1987)($1.5m award for pain and suffering reduced to $700,000); Merrill v. Albany Medical Center, 512 N.Y.S.2d 519, App. Div. (1987)(brain-damaged baby's award reduced from $12.3m to $6.1m); Vialva v. City of New York, 118 App. Div.2d 701, 499 N.Y.S.2d 977 (1986)(pain and suffering award cut from $400,000 to $100,000); and Schneider v. Memorial Hospital for Cancer, 100 App. Div.2d 583, 473 N.Y.S.2d 524 (1984)(lost wages reduced from $250,000 to $100,000; husband's loss of services from $300,000 to $100,000).

before trial.[19]

Finally, in order to reduce the incidence of doctor carelessness and patient injuries, rather than just to contain patient tort claims, jury awards and doctor insurance premiums, the Legislature has instituted a number of significant measures designed to enhance the quality of care in New York hospitals. First, each hospital is required to have programs for careful evaluation of physician competence before staff privileges are granted to new doctors, for periodic reviews and upgrading of the credentials and competence of the existing staff, for identification, investigation and reporting of all adverse incidents in patient treatment, and for hearing and resolving patient grievances about their care. Second, under the authority of these statutes, the Department of Health has promulgated extensive minimum standards for hospital quality assurance programs, a set of clinical practice guidelines and protocols for departments of surgery, anesthesia, radiology, emergency medicine, and obstetrics, and regulations limiting the working hours of residents and emergency physicians.[20] Third, the Office of Professional Medical Conduct, established as part of the 1975 legislation,[21] now monitors the performance of physicians through information reported from malpractice insurers, hospital incident reporting systems, and patient complaints. This office takes responsibility for suspending or revoking licenses and for restricting clinical privileges of licensed physicians.

As the foregoing recital makes clear, New York has adopted

---

[19] Another common law development is the Court of Appeals' recent determination that loss of enjoyment of life is not an allowable element of damages where the injured patient lacks cognitive capacity to understand the loss. McDougald v. Garber, 73 N.Y, 2d 246, 538 N.Y. S. 2d 937 (1989). Legal scholars believe that the decision will slow the growth of personal injury damage awards. Kolbert, Court Curbs Awards for "Loss of Enjoyment." N.Y. Times Feb. 22, 1989; 138:B1,B4.

[20] N.Y. Comp. Codes R. Regs. tit. 10, sec 405 (1988)

[21] 1975 N.Y. Laws sec. 28

and extended a broad array of reforms to its existing medical, legal, and insurance system in what until this year had been a vain effort to stem the rising tide of malpractice suits and premiums. With this background of developments in statutory and case law, the Harvard Medical Practice Study was commissioned by the state government.

Indeed, the roots of our study extend back far earlier than the 1986 statute which mandates the Study. As an element of the 1975 legislation which responded to the first malpractice crisis in New York, then Governor Carey appointed a Special Advisory Panel to review the broader parameters of this problem. In its Report on Medical Malpractice,[22] the Panel observed that "the field of medical malpractice constitutes one of the most emotional and critical socioeconomic problems in America" (p. 9), "the uninterrupted delivery of health services to the citizens of New York cannot be assured if the medical malpractice problem is not resolved and if matters of financing the liability of physicians and hospitals and the compensation of injured patients are not placed on a stable footing" (p. 57), and "tort law changes to remedy specific situations or defects do not go to the heart of the matter and will not, therefore, prevent unacceptable cost rises and an eventual breakdown of the system" (p. 31). Accordingly, the primary recommendation of the majority of the Special Advisory Panel was that "fundamental reform of the present tort law/liability insurance system ... be undertaken ... the overriding concern should be to create a system of compensation for adverse medical outcomes resulting from medical treatment, whether or not caused by negligence ... [together with a] companion effort ... strengthening quality control over medical practice ... to be sure that proper care has been rendered" (p. 57).

---

[22] McGill, Chair., Report of the Special Advisory Panel on Medical Malpractice, State of New York (1976)

Within the deliberations and the report of the Special Advisory Panel, however, significant gaps were acknowledged in our knowledge and understanding of how the current tort litigation and liability insurance systems were operating and about the problems that any alternative compensation and prevention program would encounter. Before undertaking a fundamental transformation of the present regime, that information gap would have to be closed. We have witnessed 15 years of soaring malpractice claims and premiums and a host of legislative measures in every state of the nation. Remarkably, however, little systematic research has been conducted to determine the dimensions or epidemiology of accidental medical injuries to patients, the degree of disability, the loss and compensation secured by patients from all sources, tort and non-tort, and the effectiveness of either tort liability or medical discipline in achieving more careful treatment by doctors and other providers. That is why in 1986, the New York Legislature finally mandated a systematic analysis "of the extent of medical injury ..., the extent to which such injuries are compensated ... and the feasibility of medical compensation systems that would compensate victims irrespective of fault."[23] This Report of the Harvard Medical Practice Study is our response to that mandate, and happily, the somewhat calmer malpractice atmosphere in 1989 will permit more considered reflection on our findings.

[23] 1986 N.Y. Laws ch. 266, sec. 39 (McKinney 1986)

Chapter 2

THE POLICY DEBATE ABOUT MEDICAL MALPRACTICE

I. Introduction

The major emphasis in our research effort as described in this Report was upon an empirical investigation of what New York's medical and legal systems are actually doing with respect to medical injuries rather than upon a normative evaluation of what they should be doing. For the last two decades this state and the country as a whole have been well supplied with often heated commentary of that latter type, usually from starkly opposed perspectives. Indeed, within our own multidisciplinary team of doctors, economists, statisticians, and lawyers, a considerable range of views are held about the best way to respond to the human problems with which we have been grappling. However, we decided that our comparative advantage consists of the ability to do serious research and analysis of the crucial factual assumptions underlying the policy debate, and in that regard we have been able to reach a firm consensus about what our investigations reveal.

At the same time, we recognize that empirical research about any area of public policy cannot be conducted in a vacuum: it must be connected up to the policy controversies that need factual illumination. Thus, as a prelude to our inquiry, we had to immerse ourselves in the current debate about medical malpractice law and to understand the sharply contrasting views expressed about the tort system and possible alternatives. We found a profound division about these issues not only in the political arena but also within the scholarly literature, and not only about medical accidents but about the broader field of personal injuries from consumer products, motor vehicles, the workplace and elsewhere. Thus, it is useful at this point in the Report for us to state as succinctly and as fairly as we can the nature of that debate in order to make clear why we made the

crucial research decisions that we did.[1]

II. The Tort System

Medical practice and injuries are governed by certain long established principles of the common law tort system, which also define liability for motor vehicle and consumer product (though not workplace) injuries.

(i). When a person is injured as a result of the wrongful behavior of another (e.g., a physician or other medical care provider),[2] then the victim (i.e., the patient) is entitled to recover for all losses -- both financial and nonpecuniary -- caused by such fault.

(ii). In the absence of culpable behavior, the doctor is not legally responsible for injuries suffered by his or her patient: instead, such losses must be borne by the victim personally or by the broader community through its various programs of public and private loss insurance.

(iii). Disputes about whether an instance of medical treatment was in fact careless and about what injuries the victim did suffer as a result are ultimately resolvable in a civil trial before a jury, although in practice some 90% of such claims are settled by the parties and their lawyers through voluntary negotiation before a trial.

(iv). If and when some legal fault and liability are established through this process, the compensation received by the victim almost invariably will be paid not by the individual who was careless but rather by a liability insurer for an independently practicing doctor or by the institution (and/or its

---

[1] With some expansion and clarification, this chapter draws upon ideas expressed in our 1987 Report of the Pilot Study. A much larger treatment of this subject, with extensive references to the literature, is contained in a document entitled Legal Policy for Medical Injuries (1988), a background paper written for the American Law Institute's Personal Injury Project by Paul Weiler, the senior lawyer member of this Medical Practice Study Group.

[2] In the medical field, the source of a medical accident may be not only the doctor treating the patient but also the nurse or the many other actors within the health care system and also the variety of organizational procedures that are not attributable to any one individual. For convenience sake in this Report, however, we generally use doctor as a shorthand reference to all such providers, personal or organizational, whose medical management may contribute to patient injuries.

insurer) by whom the doctor or other provider in question was employed.

This system of fault-based tort liability has traditionally been seen as securing two distinct though generally compatible objectives. First, the values of corrective justice and fairness require that the losses that have already occurred as a result of careless action should be shifted (as far as is possible) from the shoulders of the innocent patient-victim and made the responsibility of the culpable party.[3] Second, the prospect of being sued and having to pay for such losses will serve as an incentive (both financial and emotional) to doctors to provide more careful and safer treatment to their patients in the future.[4] The flip side of this policy rationale for the fault system is that if the patient injury has been produced as a purely accidental byproduct of careful medical treatment, imposing liability is not justified. To do so would unfairly single out innocent doctors (rather than anyone else in the community) to be financially responsible for the plight of their injured patients, and could not deter substandard care since the latter was not the predicate for imposing liability here.

The foregoing is, we believe, a fair statement of the core assumptions of the common law of tort-fault liability. Among legal scholars, however, the corrective justice rationale is no longer in general favor. A principal reason is that with the emergence of widespread (even compulsory) liability insurance, the careless doctor no longer actually pays the tort bill for the patient's injuries. Instead, the individual doctor's insurer pays any tort award with funds that are accumulated through liability insurance premiums paid by all doctors in the relevant specialty and region. In fact, in New York the excess insurance (over $1 million per case) for physician malpractice is now

---

[3] The most systematic contemporary analysis of tort law as a mode of corrective justice is contained in the writings of Ernest Weinrib; for a recent presentation of his position, see his paper, Understanding Tort Law, 23 Val. U. L. Rev. 485 (1989).

[4] A comprehensive review of this deterrence rationale for tort litigation in the medical arena is to be found in Bell, Legislative Intrusions into the Common Law of Medical Malpractice: Thoughts About the Deterrent Effect of Tort Liability, 35 Syracuse L. Rev. 939 (1984).

provided and paid by the hospital with which the doctor is associated. In turn, doctors (and hospitals) pay for these often sizable insurance premiums through the fees charged patients for medical services: the vast majority of these doctor fees are actually paid by private or public health insurance programs that themselves are funded by premiums charged to individuals and employers or taxes collected by governments. Thus we now have an elaborate set of insurance arrangements under which healthy citizens are required, in effect, to purchase a form of disability insurance against the risk of at least some injuries arising out of their medical treatment.

The actual function of tort litigation, then, is to serve as an entry point into this socially generated disability fund and to determine which injured patients are to collect and how much. If one adopts that point of view, then traditional concerns about securing corrective justice as between an injured victim and an individually culpable actor have little play. A more pragmatic policy analysis will seem appropriate, under which one asks the following questions:

(i). How sensibly does this entire system distribute compensation to injured patients from funds accumulated from the broader constituency of the health care system?

(ii). How economic is administration of this disability insurance program in money, time, and emotional demands upon the parties?

(iii). How effective is such a program for prevention of substandard medical treatment and thus for the protection of all patients from iatrogenic injuries before they might occur?

1. Compensation. As we noted above, the tort system compensates all losses of those patient-victims injured due to the fault of their doctors but compensates no losses of those injured patients who cannot establish physician culpability. From the point of view of its critics, this substantive tort

policy displays two crucial deficiencies of the system as a vehicle for compensatory insurance.[5]

The most obvious such flaw concerns the plight of those patients who recover nothing for severe injuries that create real needs for financial help to pay for additional medical care and rehabilitation and to serve as a source of income during the period when the patient is unable to work and earn a living. These needs of the injured patient are exactly the same irrespective of whether the carelessness of a doctor happened to contribute to the original incident. Thus, from this perspective it will seem arbitrary and unfair to deny this one large category of needy victims any access at all to the disability insurance fund which has been accumulated in the manner described.

More recently, tort law has been charged with a second, contrasting defect. The commitment to full compensation of those victims who do establish the fault prerequisite is said to be incompatible with sound principles of loss insurance.[6] The most apparent violation of these principles is found in the award of sizable, at-large monetary damages for such essentially nonpecuniary losses as pain and suffering or loss of companionship. A similar failing is now also found in the right to full replacement of all lost earnings or treatment costs, without any use of the deductible or coinsurance formulas which are standard in both private and public insurance.[7] While the tort regime has its own historic reasons for this dichotomy between full compensation and no compensation, the point of this present day critique is that such a substantive policy simply cannot be justified as a sensible compensation program.

---

[5] For a lengthy statement of this view about the tort system as a whole, see Sugarman, Doing Away with Tort Law, 73 Calif. L. Rev. 558 (1985), and to much the same effect about medical malpractice litigation itself, see Gellhorn, Medical Malpractice Litigation (U.S.) -- Medical Mishap Compensation (N.Z.), 73 Cornell L. Rev. 170 (1988).

[6] Recent expositions of this critical perspective on the current principles of tort damages include Danzon, Tort Reform and the Role of Government in Private Insurance Markets, 13 J. Legal Studies 517 (1984) and Priest, The Current Insurance Crisis in Modern Tort Law, 96 Yale L. J. 1521 (1987).

[7] Indeed, in those jurisdictions that, unlike New York, have not reversed the collateral source rule, the successful tort plaintiff may well recover considerably more than full compensation for certain loss items.

2. Administration. Irrespective of whether some substantive grounds can be advanced for the fault principle, it is clear that the latter does entail a substantial administrative burden for the system. To make the crucial judgment about whether the doctor was careful or careless involves a reconstruction of an event which occurred years earlier, and a contest between expert witnesses trying to educate a lay jury about how to assess what often are complex and subtle questions on the appropriate standard of medical care. Even if such judgments can be made accurately and reliably (an assumption which the medical community vigorously rejects, and about which we will provide some evidence in Chapters 6 and 7),[8] the litigation process is certainly expensive in time and money. The malpractice insurers in New York now spend well over $10,000 on average to defend every malpractice claim, meritorious or not, while successful patient-claimants pay their own lawyers a fee which can range up to one third or more of the total award. Thus, even when one ignores the cost of securing and investing the insurance funds and focuses just on the process of claims administration and distribution, not much more than 40% of the total amount expended in this claims process actually reaches the injured patients themselves as compensation for their injuries.[9]

3. Prevention. By and large, defenders of the tort system would readily concede that this program is not a particularly sensible and economic mode of insurance and compensation. In their view, however, the admitted problems created by hinging the patient's recovery on proof of the doctor's negligence are a price worth paying in the effort to deter careless medical practice by all doctors. At the same time, a third line of

---

[8] For that reason, the profession has recently proposed replacement of civil jury trials in malpractice actions with a proceeding in front of a revamped medical licensing board; see American Medical Association/Specialties Society Medical Liability Project, A Fault-Based Administrative System: A Proposed Alternative to the Civil Justice System for Resolving Medical Liability Disputes (1988). Our study will also shed some light on the reliability of doctor judgments about medical negligence.

[9] See J. Kakalik & N. Pace, COSTS AND COMPENSATION PAID IN TORT LITIGATION (1986).

argument by critics of the tort system is that the latter is not really an effective instrument with which to improve the quality of medical care.

Again, there are two aspects to this critique. One is that the actual monetary sanction which, in principle, is to be visited upon negligent doctors itself depends on the fortuitous occurrence and severity of the patient's injuries, rather than on the nature and quality of the doctor's culpability. Thus, a momentary inadvertent slip by a surgeon who is ordinarily very attentive and meticulous, but who in this case has unfortunately inflicted permanent long-term brain damage on a patient, will make this doctor liable for millions of dollars in tort damages.[10] On the other hand, another doctor whose deviation from appropriate standards of care has been both deliberate and egregious but whose patient suffers a much less serious injury, will escape with only moderate financial liability. Indeed, if the potential damages are reasonably small, there will be no liability at all, because neither the patient nor a lawyer will have sufficient incentive to make the substantial up-front investment in bringing suit against the doctor.

Of course, the real life tort litigation system does not operate in accordance with the formal legal principle that would produce such discordant results. A liability insurer actually pays all these awards, whether large or small. What doctors pay is a standardized insurance premium that depends on the nature of their specialty and geographic location. But the presence of such liability insurance in the background has attracted the even sharper criticism of this deterrent rationale for the tort system. If the doctors who are careless do not personally bear the cost of their default, then how can the prospect of a tort award actually serve as a deterrent against substandard care by others? One might, of course, adjust the size of the doctor's insurance premium in light of individual claims experience (as

[10] The tort system's assumptions about its ability to meaningfully influence and control the level of attention of individual actors is criticized in Latin, Problem-Solving Behavior and Theories of Tort Liability, 73 Calif. L. Rev. 677 (1985), and Grady, Why Are People Negligent? Technology, Nondurable Precautions, and The Medical Malpractice Explosion, 82 NW. U. L. Rev. 293 (1988), the latter piece focusing specifically on the medical accident setting.

New York State now requires).[11] However, there is considerable doubt about whether a meaningful and actuarially credible experience rating program can be devised for liability insurance against claims that are as occasional an event to the individual doctor as are malpractice suits (by contrast, e.g., with the much more extensive workers compensation experience of a large employer).[12]

Still, to suggest that liability insurance fully insulates doctors from the impact of the tort system is somewhat unrealistic. A far more significant factor than even a demerit adjustment in an insurance premium is the loss of practice time and earnings, as well as personal morale and professional reputation, when the doctor's quality of medical care is attacked through tort litigation.[13] The problem is that many of these tangible costs are visited upon doctors from the mere fact of being sued, rather than just when they have been found liable. Thus, the tort system actually inflicts its tangible, uninsured sanctions on the majority of doctors whose mode of care is eventually vindicated, but whose patients have sought to use this legal avenue to secure some financial redress for their serious injuries. A final critique of the tort system, then, is that its deterrent instrument is much too crude in its operation, and serves primarily to induce unnecessary and wasteful modes of defensive medicine designed to make it less likely that the doctor will be sued rather than more likely that the patient will be cured.[14]

---

[11] For a description and appraisal, see Darling, The Applicability of Experience Rating to Medical Malpractice Insurance, 38 Case W. Res. L. Rev. 254 (1987).

[12] For an analysis of the sophisticated experience rating program in no-fault WC insurance, see R. Victor, WORKERS COMPENSATION AND WORKPLACE SAFETY: THE NATURE OF EMPLOYER FINANCIAL INCENTIVES (1982).

[13] See, generally, Bell, note 4 supra, as well as Chapters 9 and 10 infra.

[14] For an attempt to measure the defensive (though not necessarily the wasteful) reaction of doctors to present-day malpractice litigation, see Reynolds, Rizzo & Gonzales, The Cost of Medical Professional Liability, 257 J. A. M. A. 276 (1987).

III. No-Fault Alternative

Finding fault with the tort system's inability to meet certain lofty but abstract ideals is rather easy. A much tougher challenge is devising another system that could actually do a better job of securing the variety of our often incompatible goals.

To consider fundamental alternatives to tort-fault liability (rather than just the specific reforms that New York has made in its tort system, as described in Chapter 1), one might move in either of two directions. In one direction lies the no-liability option, under which the tort system would be dismantled (perhaps through private contract)[15] and injured patients would pursue redress through the same public and private systems of loss insurance that are available to victims of any other disabling injury. Proponents of this policy tack often advocate considerably improved systems of social insurance for medical costs and lost earnings, and stiffer regulatory sanctions against risky behavior.[16] The implication of this no-liability option remains, however, that providers of medical care would shoulder no special burden to pay for the losses suffered by the patient-victims of accidents occasioned by medical care.

The opposite path would move us from fault to no-fault liability. With no-fault medical accident insurance (provided on a mandatory or a voluntary basis by private or by public carriers), the legal responsibility for adverse consequences of medical treatment would still be handled under a special program devised and, we assume, paid by the health care system. Entitlement of the patient to compensation would turn not on whether the doctor had been at fault, however, but simply on the fact that the injury had been caused by medical treatment. By analogy to no-fault workers compensation, the crucial test in such a patient compensation program would not be whether someone

---

[15] For an exploration of the pros and cons of this possibility, see Robinson, Rethinking the Allocation of Malpractice Risk Between Patients and Providers, 49 Law & Contemp. Probs. 173 (Spring, 1987), and Atiyah, Medical Malpractice and the Contract/Tort Boundary, ibid. 287.

[16] See, e.g., Sugarman, note 5 supra.

could be blamed for the injury but simply whether the injury "arose out of and in the course of" medical treatment.

The no-fault option surfaced in the debate about medical malpractice when dissatisfaction was felt in the early 1970s about the performance of the tort system.[17] New Zealand adopted a version of such categorical no-fault compensation for the victims of medical accidents as a feature of its broader social insurance approach to all other accidental injuries.[18] Sweden followed with a separate and self-contained patient compensation scheme.[19] As detailed in Chapter 1, in 1975 the members of New York's McGill Commission apparently endorsed the idea in principle as well.

For various reasons, however, significant difficulties impede the creation and implementation of no-fault patient compensation; thus this option stayed largely in the background while the debate focused on a variety of tort reforms. With the resurgence of malpractice litigation and premium costs in the 1980s, not just in the United States but elsewhere in the world,[20] the no-fault idea is now back in vogue. Finland has adopted a program largely modeled on Sweden's,[21] the British Medical Association has expressed interest in this idea,[22] and a government commission in Canada has just proposed adoption of no-fault there.[23] Meanwhile, in this country the states of Virginia and Florida have recently adopted no-fault compensation schemes for infants who suffer severe and permanent brain damage

---

[17] See, for example, Havighurst & Tancredi, Medical Adversity Insurance -- A No-Fault Approach to Medical Malpractice and Quality Assurance, 613 Ins. L. J. 69 (1974), and O'Connell, No-Fault Insurance for Injuries Arising from Medical Treatment; A Proposal for Elective Coverage, 24 Emory L. J. 21 (1975).

[18] For a detailed description of the New Zealand system written for an American audience, see Gellhorn, note 5 supra.

[19] Specifics of the design and operation of the Swedish program are set out in M. Rosenthal, DEALING WITH MEDICAL MALPRACTICE: THE BRITISH AND SWEDISH EXPERIENCE 131-206 (1988), and in Oldertz, The Swedish Patient Insurance System -- Eight Years of Experience, 52 Med.-Legal J. 43 (1983).

[20] For the comparable trends in Canada, see D. Dewees, P. Coyte & M. Trebilcock, Canadian Medical Malpractice Liability; An Empirical Analysis of Recent Trends (1989), and in the United Kingdom, see C. Ham, R. Dingwall, P. Fenn & D. Harris, Medical Negligence: Compensation and Accountability (1988).

[21] See Braham, No-Fault Compensation Finnish Style, Lancet 733 (Sept. 24, 1988).

[22] See British Medical Association, No-Fault Compensation Working Party Report (1987).

[23] See R. Prichard, Medical Malpractice (1989).

during obstetrical delivery, a major sore spot within the present medical malpractice system.[24] In this setting, New York commissioned this study of the respective values of the no-fault and tort-fault approaches to medical accidents.

When one reads the literature about the no-fault alternative, it turns out that this model does have some significant comparative advantages over tort/fault, but also has considerable problems of its own. On the positive side, the virtues of no-fault are particularly responsive to the deficiencies we saw in the tort system.[25]

1. Compensation. The scope and coverage of this disability insurance would be broadened to cover the victims of all medical injuries, not just those "lucky" enough to be able to prove that their injuries happened to result from the negligence of a doctor or some other financially responsible provider. At the same time, at least if one were to emulate the philosophy (though not necessarily the specifics) of the no-fault model in the workplace, the amount of compensation paid to any one victim would likely be scaled downwards (particularly with respect to pain and suffering, but also with respect to some proportion of lost earnings). This would be done partly to free up some of the funds needed to finance the expanded coverage for all victims, but also because (for the reasons alluded to earlier) such a compensation format is more in accord with sound insurance principles.

2. Administration. Elimination of the legal contest about the crucial tort issue of whether a doctor was at fault would remove one major source of litigation. There would still be disputes about the patient's entitlement to any and how much compensation, but these would focus on the more technical

---

[24] For a detailed description of the background to and the design of the Virginia program, upon which the Florida provision is largely modeled, see White, Innovative No-Fault Tort Reform for an Endangered Specialty, 74 Vi. L. Rev. 1487 (1988).

[25] A useful exposition of the case that can be made for medical no-fault is to be found in Gellhorn, note 5 supra.

questions of the precise source and dimensions of the patient's injury, rather than on whether the doctor was to blame for what happened. This change would spare both doctor and patient the emotional and financial costs of such a heated legal conflict and permit the community to shift much of the resources now spent on lawyers, expert witnesses and court personnel to direct reimbursement of the patients themselves for the losses they have suffered.

3. Prevention. Of course, these compensatory and administrative values would be served even better if one also dispensed with the requirement of "cause" as a predicate for patient recovery, and adopted instead a broader social insurance approach to these patient needs.[26] From this perspective it might seem strange to establish a program whereby patients are guaranteed insurance compensation for the health care costs from disabling injuries they suffer inside the hospital, when there is no such guarantee of health insurance coverage for the injuries or illnesses that brought the patient into the hospital in the first place. However, there is an important reason why one might want to adopt a specific no-fault program for a separate category of injuries caused by an activity such as health care, even though such a categorical program does entail significant compromises in the easy and evenhanded compensation of past injuries. Such a program would keep within the medical care system itself the legal responsibility for those injuries which flow out of its own operations, in order to maintain a significant financial incentive within that system for safer care and better prevention of future injuries to all its patients.

This is the crucial difference between the model of no-liability even in the case of fault (to which we referred earlier), and the model of full liability irrespective of fault considered here. Under a no-fault regime, the doctor (more likely, the hospital) is subject to essentially the same legal

---

[26] This is the gist of the argument of Sugarman, note 5 supra, against the no-fault as well as the tort/fault models of liability as a system of victim compensation.

and financial responsibility as under tort law for those injuries which could and should have been avoided by more careful treatment (i.e., the fault-caused injuries). Thus, to the extent that the prospect of such liability does furnish some meaningful incentive to providers to take greater care, such a legal incentive would be largely untouched by the fact that such awards would now be paid under a no-fault format. What no-fault does is to establish a new and broader liability on the medical care system to pay for patient losses from all those iatrogenic injuries which might not now seem avoidable by currently feasible methods. This additional legal responsibility would build right into the health care system a significant incentive to develop and install innovative quality assurance techniques and equipment that eventually would make feasible the avoidance of more and more of the medical accident toll.

* * * * * * *

Given these apparent advantages of no-fault, the question one might ask is why this model has not been taken very seriously in the medical arena as the burdens of the tort liability regime have seemingly become more and more oppressive. One answer is that even those people who are generally sympathetic to no-fault (e.g., for motor vehicle accidents) have several reservations about the viability of such a program for medical injuries.[27]

1. Cost. An immediate worry is the potential financial cost of having to compensate all patient injuries within a no-fault scheme. Modern medical care is an inherently risky enterprise. The major precursor to our own empirical study, the Medical Insurance Feasibility Study in California in the mid-1970s,[28]

---

[27] See, for example, Keeton, Compensation for Medical Accidents, 121 U. Penn. L. Rev. 590 (1973), and Calabresi, The Problem of Malpractice: Trying to Round Out the Circle, 27 U. Toronto L. J. 131 (1977). A succinct statement of the standard present-day arguments against medical no-fault is presented in P. Danzon, Medical Malpractice: Theory, Evidence and Public Policy 152-58 (1985).

[28] California Medical Association, Medical Insurance Feasibility Study (1977). The nature and results of this pioneering research are described in more detail in Chapter 3 of this report.

found the risk of disabling injury to patients to be roughly one in twenty hospital admissions. However, only a small fraction of these injuries produced actual tort recoveries under the fault principle. In theory, all such iatrogenic injuries would be eligible for compensation under a no-fault scheme. Thus, because the primary motivation for reform efforts in the malpractice arena has been to stem spiraling liability insurance premiums, adoption of a no-fault program which would add a host of new claims for compensation has seemed precisely the opposite direction in which to go.

2. Cause. While the substitution of cause for fault as the precondition to compensation does broaden the potential range of patient recovery, it still places an intrinsic limit on those injuries which are to be the financial responsibility of such a patient compensation program. Administration of this new borderline is a second source of serious concern.

We noted earlier that under no-fault workers compensation, the standard inquiry is whether the worker's injury "arose out of and in the course of employment," and in the case of workplace accidents use of this formula has occasioned little practical difficulty.[29] Unfortunately, no such simple formula will work in the medical area.

By comparison with workplace accidents in which the employee typically goes to work healthy and leaves hurt, the patient who goes to a doctor or hospital ordinarily is already in an unhealthy condition, often requiring medical intervention, which itself can be traumatic in character. Severe difficulties could then arise in distinguishing those disabling injuries that actually were the adverse consequence of medical care (e.g., an infection picked up as a result of conditions in the operating room) from those that were attributable to the original illness (e.g., an infection that was due to the cause /or site of the original wound). And even if the patient's disability is clearly

---

[29] Conley & Noble, Workers Compensation Reform: Challenge for the 80s, 1 Research Report of the Interdepartmental Workers Compensation Task Force 1, 57 (1979)

attributable to the treatment, the patient would not necessarily have a right to compensation because some such traumatic effects may be the inevitable byproduct of treatment that was in the patient's broader interests; consider, for example, the amputation of a cancerous limb or organ. In light of examples such as these, most commentators have been dubious that one could regularly disentangle the precise (and compensable) medical contribution from the many other factors that may contribute to a patient's ultimate disability, at least with any greater ease and economy than the tort system does now with fault. Moreover, the number of claims for which such a determination would have to be made would be considerably greater. As a result, the anticipated savings in administration and litigation expenses under a no-fault program might be nowhere near enough to help pay for the latter's broader scope of compensation.

3. Care. A final concern about the adoption of no-fault in the medical arena is that this would unduly dilute the necessary incentives for safe treatment. As we saw earlier, a no-fault scheme does place on the health care system a legal responsibility for medical injuries, and thus does maintain a significant financial incentive for the prevention of those injuries which can be avoided by reasonable efforts.[30] However, there are a number of important differences in the manner in which the tort and the no-fault liability systems function in this respect.

(i). While the overall financial costs of no-fault might well be higher because it must compensate all injured patients, the amount paid to the minority who are the victims of careless treatment would likely be less than under the full compensation principle of the tort regime. Thus, a smaller financial incentive would still be felt to avoid this subset of negligent incidents (to the extent these could actually be isolated within

[30] K. Viscusi & M. Moore, Compensation Mechanisms for Job Risks: Wages, Workers Compensation, and Product Liability 151-78 (1989) estimate that the incentives generated by the present day, experience rated workers compensation programs have spurred American employers to additional safety precautions that have reduced workplace fatality rates by more than 25% from what they would otherwise be.

the broader universe of cases facing the provider).

(ii). Again, if the workplace example were to be followed, no-fault medical accident insurance would more likely be provided and paid for by organizations, such as hospitals, clinics and HMOs, rather than by the individual doctors who provide the treatment and whose personal carelessness may be the actual cause of a patient's injury.

(iii). Irrespective of who actually pays for the patient's losses (the liability insurer does so under the current tort system), a bigger difference in no-fault is that there would be no opportunity through the jury trial to let the outside public scrutinize the behavior of doctors and hospital staff, and perhaps to condemn those practices that are inappropriate. To the extent, then, that the influence of the fault-based tort system is traceable more to subtle factors of moral psychology than to direct pecuniary expenditures, much of the former influence would be lost under no-fault.

(iv). Another concern often expressed about no-fault is that it also loses sight of the responsibility and culpability of the individual victim in the occurrence of many accidents. Although the patient in the hospital is typically a more passive factor in medical accidents than is the employee in the workplace or the consumer-user in a product-related case, a considerable number of medical injuries (e.g., from prescription drugs) could have been avoided by greater patient care and attention to the doctor's instructions. The worry about no-fault is that a guarantee of compensation to the patient (however he or she may be hurt) will reduce the latter's own instincts for self-preservation (including, inter alia, the risks of aggressive medical treatment), and thus add to the injury toll within the hospital and the cost burden upon the patient compensation scheme.

(v). Finally, just as many critics lament the tendency of the tort system to induce wasteful modes of defensive medicine by the individual physician seeking to erect a barrier to litigation, so also one would anticipate an analogous unhappy reaction on the part of the hospital if the latter were made financially responsible for a no-fault compensation scheme for

the benefit of its patients. We shall see in Chapter 6 that patients are by no means homogeneous with respect to the risk of iatrogenic injury. Some hospitals may be tempted to reduce their financial exposure to such insurance costs by refusing access to their facilities and care to those types of patients who are most likely to experience adverse events. That reaction would be of the same genre as the patient dumping which is said to have taken place within the incentive structure of Medicare's new Prospective Payment System.[31]

The general theme which runs through these several concerns is that medical no-fault would inevitably function in too unobtrusive and impersonal a fashion, focusing just on the financial responsibility of the enterprise. Such a system would be nowhere near as successful as the morality play of tort litigation, with its focus on the personal blame and accountability of human beings, in galvanizing the health care system into serious efforts to eliminate substandard care.

IV. The Research Agenda

The foregoing is just a brief sketch of the arguments that are made for and against these two contrasting liability systems for dealing with medical injuries. Experts in the subject will appreciate the necessary qualifications and available responses to the variety of opposing claims we have set out, and also realize that any real life program, whether fault or no-fault, can be refined to accommodate at least some of these concerns. Our objective here is not to enter directly into this policy debate, nor to commit ourselves in favor of one position or another: as we stated earlier, there are considerable differences of view within our own group about these policy alternatives. We simply want to underscore that real issues and arguments exist on both sides of this debate. Further, lurking beneath the surface of the polarized positions are some crucial factual assumptions

[31] J.F. Newhouse, Do Unprofitable Patients Face Access Problems? Health Care Financing Review 11 (1989)

-- about the incidence of medical injuries and carelessness, about the needs of injured patients for compensation, and about the legal incentives for more careful medical treatment. These assumptions often have little or no factual basis and, as we shall see, may be misguided. The aim of our research is to focus on and to investigate each of these key empirical dimensions of the policy problem.

The details of our research efforts are described in succeeding chapters. To state them in summary form here, we have studied the hospital records of a state-wide sample of roughly 30,000 patients in New York hospitals in 1984. Through a careful review of these records, we have determined which of these admissions produced adverse consequences from the medical services rendered, which of such consequences were due to the fault of a doctor or other provider, and where it occurred, the degree of such fault (ranging from minimal to severe).

After determining which hospital admissions involved iatrogenic injuries, we then conducted a survey of the affected patients to find out the financial and personal consequences to them of such injuries (as compared to a control group of patients who were not injured). What kinds of additional medical costs were incurred? What extra losses of earnings were experienced? What impairments were suffered in the ability to function in the household and in regular non-work activities? How many of these losses were actually made up through the variety of existing programs, private as well as public, tort as well as non-tort?

A third area of inquiry sought to assess the response of the tort system to these events, with respect to the doctor as well as the patient. What was the level and distribution of tort claims arising from our sample of cases, whether from those we judged to be or not to be negligent adverse events? What consequences, financial and psychological, do tort proceedings tend to have on a doctor personally (as opposed to the effects on the doctor's insurer), and what are the apparent consequences to the doctor's subsequent practice? And finally, can one detect through systematic econometric analysis whether the considerable

variation in the risk faced by New York doctors in the different regions of the state had any discernible impact on reducing the rate of injury, or of careless injury, in our overall sample of 30,000 admissions in that year?

Each component of the overall research project will produce data important for a variety of perspectives on the medical injury problem. Thus our systematic review of a large sample of hospital admissions enabled us to develop a comprehensive epidemiological picture of iatrogenic injuries: portraying the kinds of patients and disease conditions, types of doctors, hospitals and providers, or variety of medical procedures and settings that are especially accident-prone. Such material is crucial for those who are responsible for quality assurance within the medical system, indicating to them where they should place their priorities and giving them clues to the underlying causes of such accident patterns. It does little good for tort litigation or other external legal regimes to escalate the legal sanctions against substandard care if those who are actually practicing within the health care system do not have and cannot obtain the information and understanding which is necessary to devise sensible practice protocols and safeguards that can produce a significant reduction in adverse outcomes.

But over and above the contribution this study can make to the direct enhancement of quality of care within the medical system, our research is designed to shed light on the key issues we have just seen dividing the proponents of fault and no-fault liability.

1. Affordability. As stated earlier, in considering whether no-fault coverage is viable for medical accidents, the immediate question is whether it would be financially affordable. It should be recognized that, from a more fundamental perspective, this is a spurious question. Once the medical injury has occurred and losses are suffered by the patient, the losses must be borne and thus "afforded" somewhere, either by the patient personally or by some other segment of the community. Still, if someone is considering whether to alter the nature of

the health care system's own responsibility for such injuries, they will want some guidance about the cost of the insurance that would be needed. And as we said, because the earlier California Medical Insurance Feasibility Study found a substantial incidence of iatrogenic injuries, but tort recovery by only a small fraction of these victims, the often voiced objection to no-fault patient compensation is that it would entail far too steep an increase in already burdensome malpractice insurance premiums.

However, one cannot estimate the potential costs of a no-fault program simply by knowing how many patients were injured and what proportion was due to negligence or was purely accidental. One also needs to know the size of the losses those injuries actually caused patients, particularly losses of the type a no-fault program would be likely to compensate. The extent of a disabled patient's earnings loss, for example, depends not only on the occurrence but also on the severity of the physical impairment and, even more, on the interaction of this impairment with the age, occupation and other personal characteristics of the victim.[32] It is possible that a large share of the more costly accidents is actually caused by negligent treatment -- cases which are potentially litigable under the tort system -- and so the gap between total patient losses and those which would be potentially compensable under no-fault may be much smaller. The major aim of the patient survey, then, is to learn from the patients themselves what losses they have suffered -- not just financial in character, but also in the enjoyment of their daily lives -- and how such losses are distributed between fault and no-fault accidents.

This is still just one side of the equation: one also needs to know how much of these losses the program needs to compensate. Suppose one assumes that no-fault patient insurance would be a secondary source of compensation, serving as a backstop to the general medical disability insurance provided by the variety of private and public programs (a role that medical malpractice

---

[32] This has been the consistent finding of a growing body of research about the economic effects of permanently disabling injuries to employees in the workplace: see, generally, M. Berkowitz & J. Burton, Jr., Permanent Disability Benefits in Workers Compensation (1987).

insurance itself now plays in New York, following the reversal of the collateral source rule). Estimates of the cost of such insurance require knowledge, then, not just of the losses the patients have suffered, but of what alternative sources of insurance coverage they had and how much of their losses was made up from these primary sources.

Nor should one assume that all these net losses must be compensated in full. From a purely insurance perspective it would make some sense to exclude the numerous shorter term disabling injuries on the assumption that these relatively modest losses can and should be handled from the victim's personal resources. Priority in any insurance program should always be given to the longer lasting disabilities that affect far fewer patients but inflict severe or even catastrophic losses on the individual and family concerned.[33] In other words, it is possible to design a no- fault program so as to target its benefits and expenditures towards the more needy cases, which should have first claim on the resources that had been made available. But if so, one can make realistic estimates of the potential costs of different versions of a patient insurance program only with detailed information from the survey of patients about the distribution as well as the aggregate size of the net losses that they have suffered from their disabling medical injuries.

2. Administrability. A second concern, which is felt not just about the tort system but also about a possible no-fault substitute, is that each such regime requires too difficult a judgment about their respective conditions for patient entitlement to compensation. Unquestionably the tort system does expend a large amount of emotion as well as money in its adversarial contest about whether the doctor was at fault as a

[33] A latent effect of the current method of paying for tort litigation through the contingent fee is that the tort system also screens out the bulk of smaller patient injuries and losses because these cases cannot produce an award amount, and thence a legal fee, that would warrant the up front costs and risks involved in filing and pursuing a tort claim.

predicate to awarding the injured patient any redress.[34] By contrast, the no-fault model in workers compensation expends less than 20% of its claims dollar on administration, or roughly one third the proportion spent in medical malpractice.[35] It is likely, however, that any attempt to transplant the no-fault model into the medical area would encounter a greater administrative burden in its causal inquiry because of the difficulty in disentangling those adverse consequences that are the accidental byproduct of medical treatment from the disabling effects of the initial illness or of its intended treatment procedure.

In actual fact, those who have reported on the experience with patient compensation plans in Sweden and New Zealand do not find this to be as great a problem in practice as one might anticipate in principle. It is true that a number of conceptual and pragmatic difficulties have been encountered in those countries which have gradually been resolved through a detailed jurisprudence. However, the reports we have read indicate that these systems have operated reasonably smoothly, with the doctors able to help their injured patients to collect the benefits to which they are entitled rather than having to defend their professional reputation against allegations of negligence.[36]

However, the only way to be confident about this issue in the New York context is actually to look at the cases here. Thus, in our systematic survey of adverse events and negligence, we instructed our reviewers not simply to give us a bottom-line judgment about whether there was "cause" or "fault" displayed in the medical records, but also to tell us the degree of difficulty

---

[34] Nearly 60% of the medical malpractice claims dollar is now expended on administration in order to get just 40% into the injured patient's hands: see J. Kakalik & N. Pace, note 9 supra.

[35] This estimate of the administrative cost share of the WC claims dollar can be calculated from the figures presented by Price, Workers Compensation Programs in the 1970s, 42 Soc. Sec. Bull. 3 (1979) on the administrative costs of employer insurers and public agencies, and from those presented in M. Bernstein, Litigation, Representation and Claimant Protection in Workers Compensation, 4 Research Report of the Interdepartmental Workers Compensation Task Force 115, 133-39 (1979), about injured worker expenditures on lawyers. These data for WC insurance costs, like those for tort liability insurance, do not include the business cost of selling the insurance, collecting and investing the premiums, and so on, and focus instead just on the costs of dispute resolution in distributing among victims the insurance funds thus raised.

[36] See Gellhorn, note 5 supra, and M. Rosenthal, note 19 supra.

that they encountered in making those judgments. We would then be able to make firmer estimates not just of the existence of this problem, but also about its dimensions (i.e., how many close calls had to be made about the causation of adverse events), and not just in absolute terms, but by comparison with the degree of difficulty in judging the negligence of the doctor. As well, a large scale sample of such cases enables us to determine whether these threshold cases tend to occur at random or in recurring patterns that would enable a policymaker to specify certain formulae for designating an event as adverse and compensable or not.

3. Controllability. Whatever the verdict about the affordability and the administrability of different programs for compensating past medical injuries, an equal if not greater policy concern is whether and to what extent these programs would control and reduce the incidence of such injuries in the future. We cannot produce empirical evidence about that role for no-fault patient compensation because no such plan now exists in New York (as we noted earlier, however, a substantial effect has been found for no-fault workers compensation in reducing workplace injuries).[37] We did set out to gather some systematic evidence about the preventive effect of present day tort litigation on medical accidents in the state because this has to be the major rationale and justification of that regime.

Four aspects of our research program are pertinent to the prevention issue. The medical records survey discloses whether the primary source of iatrogenic injuries generally, and of the more seriously disabling injuries in particular, is provider negligence or is the purely accidental byproduct of medical treatment. As well, the survey attempts to make more specific judgments about the gravity of the carelessness displayed in

[37] See K. Viscusi & M. Moore, note 30 supra. As well, the Swedish no-fault patient compensation scheme has recently turned its attention to the issue of prevention. Its entire body of case claims and payments are now being mined as a data base for research that isolates certain settings and procedures that are especially prone to iatrogenic injuries and then devises alternative, less-risky practice protocols that are reported in Swedish medical journals; see M. Rosenthal, note 19 supra, at 184-86.

these cases and its possible amenability to legal influence. Next, the cross-sectional analysis of tort claims sets out to match the responsiveness of the tort process to the judgments we have made about negligence and adverse events in our sample. How often were tort suits brought by the patients and what was the gap between potential and actual tort recoveries? How well does the distribution of such claims correlate with the seriousness of doctor negligence as compared with the severity of the patient's disability and needs? Then, the survey of doctors was intended to cast light on how physicians perceive and react to the experience or to the threat of being sued: is the signal and the sanction emitted by the legal system likely to elicit higher quality medical care on their part?

Despite these indices, the bottom-line test of the preventive impact of the tort system is found in the actual incidence and distribution of patient injuries rather than in reports of physician perceptions. Our data base of a large state-wide sample of adverse and negligent adverse events makes possible the first ever econometric study of whether the tort system -- here, medical malpractice litigation -- actually reduces such doctor malpractice and patient injuries. All the cases in our sample were formally subject to the same tort law prevailing in New York at that time: thus we could not test directly the deterrent influence of that law by comparing the performance of one group that is subject to such liability and another control group that is not. On the other hand, the real life intensity of tort litigation (i.e., the actual threat the system poses of suits and awards against doctors and hospitals) is several times greater in some parts of the state than in others. Thus, as a proxy for the either/or, tort/no-tort possibility, we have used the more-tort/less-tort variation across different hospitals and medical staffs to determine whether this variation produced any difference in the observed doctor negligence and patient injuries in the 30,000 cases we reviewed from across the state (controlling for other relevant characteristics of the patient, the illness, the hospital

facilities, and so on). With some qualifications, the results of this study should give us for the first time a more informed basis for judging whether the defenders of the tort system are correct in supposing that this regime secures sufficient benefits in quality control as to outweigh the limitations of fault-based liability as a vehicle for compensating prior victims.

V. Conclusion

The aim of our research program has been to inform the policy debate now going on in New York and elsewhere about how society can best deal with its medical accidents and malpractice. In order that our research make a meaningful contribution to this controversial subject, we have had to understand and isolate the key issues and assumptions that divide the protagonists of both the tort system and its no-fault alternatives. We have not prejudged the feasibility of any such no-fault program for injured patients, nor have we endorsed the criticisms that are made about present day malpractice litigation. Our aim has simply been to provide the relevant empirical data that would permit more informed judgments about the variety of charges and counter charges regularly exchanged by participants in the current controversy.

Chapter 3

IMPLEMENTING THE HARVARD MEDICAL PRACTICE STUDY

I. Introduction

The Harvard Medical Practice Study was designed to evaluate the function of the tort system as it deals with medical malpractice and to consider the possible effects of alternative approaches. After passage of the malpractice reforms of 1986 in New York, David Axelrod, M.D., the New York State Commissioner of Health, expressed an interest in having the Study group carry out its research program in the state. In September 1986 members of the Study met in Albany with Commissioner Axelrod and members of the Department of Health. Shortly thereafter, the Department contracted with Harvard University to undertake a series of empirical studies.

The limitations in existing data on the malpractice problem made it necessary to develop and carry out a comprehensive research program. Particularly, we required information about the incidence of injuries, both negligent and non-negligent, caused by medical management of patients. Equally important, we needed to evaluate the performance of the current fault-based tort system, including its compensation of victims and its effects in discouraging substandard care on the part of providers.

Obtaining such information required access to hospital records, to both patients who had suffered medical injuries and control patients who had not, to physicians, other providers and administrators, and to insurance data. All of this work had to be conducted with strict confidentiality. Because many of the questions that confronted us had not previously been asked, we developed and tested new instruments for several parts of our work. The fields encompassed not only medicine and law, but also economics, statistics, survey research, public policy, and sociology. As in any complex multidisciplinary inquiry, the eventual conclusions and recommendations could be no stronger than the weakest component of our work. Therefore, our team had to include people highly qualified in each of the areas involved.

This chapter describes some of the problems that arose as we undertook our research program and how those problems were managed.

II. Pilot Study and Intermediate Pilot Study

Under our first contract with the State of New York, we conducted a Pilot Study from November 1986 through March 1987 in three New York hospitals. The principal aim of the Pilot Study was to develop and test methods that would facilitate investigation of several aspects of the medical injury problem:

(i) To determine the incidence and pattern of injuries caused by medical management through a detailed review of hospital records;

(ii) To estimate the financial consequences of such injuries for the patient-victims through an in-depth telephone survey of these patients; and

(iii) To investigate the impact on doctors and other providers of involvement in such injuries.

The Pilot Study involved 2,800 patient records in three hospitals and just under 300 patient interviews. The results of the Pilot Study were reported as scheduled in a meeting in Albany in March 1987 and were released in a report to the State of New York the following month.

After completion of the Pilot Study in New York and prior to initiation of the main Study, we undertook work in the Boston area to evaluate further our methodology. Of special importance was an analysis of the reliability and validity of our process for reviewing hospital records. We were particularly interested in the potential and limitations of the hospital record as a means of revealing poor quality care. This work is cited in this report as the Intermediate Pilot Study.

III. Main Study

We began by designing a sampling method for choosing medical records, by refining the medical record review process, and by

recruiting and training medical record administrators, physician-reviewers, and other staff for the Study. During this period, we also briefed representatives of the New York State Medical Society and the local chapters of the American College of Physicians, the American College of Surgeons, the Ad Hoc Committee on Medical Malpractice of the Association of the Bar of the City of New York, and the American College of Obstetricians and Gynecologists on the purpose of the Study. In order to move ahead as rapidly as possible, much of the preparatory effort began before the formal contract was signed.

A. Some Key Events

The following are some key dates in the progress of the Study.

7/28/86	Initial meeting with Dr. Axelrod
8/25/86	Pilot Study proposal submitted to New York
9/11/86	Pilot phase began
11/5/86	Contract signed
4/1/87	Pilot phase completed
6/16/87	Sample of 51 participating hospitals selected and list sent to Health Department
7/30/87	Commissioner Axelrod sent letter to sample hospitals requesting their participation in Study
8/31/87	Harvard University and New York Department of Health agreed to contract language
9/30/87	Proposed date for completion of solicitation and approval of hospitals by the Department of Health
12/28/87	Study given approval to contact hospitals
1/6/88	Study mailed 23 introductory letters to hospital administrators
1/22/88	Contract approved by New York State Comptroller General
2/2/88	Harvard University received contract with all necessary New York approvals
3/31/88	Interim report submitted to New York
4/18/88	Hospital record review began
5/31/88	Study obtained consent from last hospital in sample
9/26/88	Patient interviews began
12/27/88	First physician interviews
4/26/89	Interim report submitted to New York

5/20/89	Hospital record reviews completed
6/ 7/89	Physician interviews completed
8/18/89	Collection of malpractice claims data suspended
8/31/89	Patient interviews completed
9/ 7/89	Discussion of some results with Dr. Axelrod
10/12/89	Presentations of some results to representatives of New York Department of Health, New York Legislature, Governor's Office, New York State Medical Society, and Hospital Association of New York State

Delays in obtaining participation of a very few hospitals presented special problems, for virtually all aspects of the work depended on the outcome of the hospital record review. Because two hospitals did not agree to join the Study until ten months after they were first approached and others were even later in providing some of their records, the record review was not completed until late May 1989. Further, the insurance claims data base was not fully usable until late September 1989. As a result, important derivative data collections were not completed until December 1989. Some analyses of our data will require additional months, and supplemental reports will likely be submitted later in 1990.

As noted in Chapter 2, we undertook to answer the following questions:

1. What are the dimensions of the problem of medical injury among hospitalized patients and what fraction of injuries results from negligence?
2. What is the relationship of medical injury to claims and litigation?
3. What are the costs of medical injury to the patient and to society?
4. What are the effects of variations in the threat of litigation on the performance of physicians and other providers?

Each of these major questions required collection and analysis of large data bases. Each posed challenging problems. The rest of this chapter describes many of the problems and how they were resolved.

B. Hospital Record Review

Key to our entire Study are reliable data on the annual incidence of medical injury and of the fraction of injury resulting from negligence. We wanted to study a period that was sufficiently recent to resemble the current situation as closely as possible but sufficiently distant so as not to compromise access to hospitals and records due to concern about self-incrimination on the part of hospitals and physicians. Further, enough time had to have elapsed after the injury so that the consequent economic losses would have stabilized and the sources of compensation would be manifest. For these reasons we decided to direct our inquiry to injuries that occurred during 1984.

1. Defining the population at risk. In principle, any encounter between a patient and a health care provider, ranging from a visit for a diagnostic test to an emergency room visit or hospitalization, can lead to unintended injury to the patient. Our focus is on what we call adverse events, or unintended injuries sufficiently serious to lead to temporary or permanent impairment or disability in the patient. Our goal is to estimate the risk that medical management will lead to adverse events. This section presents our approaches to defining adverse events, negligence and the population at risk and to determining the incidence of adverse events.

Several practical issues played key roles in determining the population at risk we would study. First, a random sample was essential for an accurate description because the potential for bias was serious if we relied solely on a convenience sample. In addition, we wanted to sample selected encounters at higher rates because of their special importance to our studies of economic consequences and deterrence. A random sample requires a centralized record keeping system delineating all possible patient encounters, which then can be used as a sampling frame. Even though every participant in health care delivery keeps records, the only centralized and uniform record-keeping system is for hospitalizations, maintained in New York as the Statewide

Planning and Research Cooperative System (SPARCS), maintained by the state's Health Department.

Second, the yield of adverse events per records sampled had to be maximized in order to make optimal use of Study resources. Clearly the risk of serious injury depends strongly on the nature of the encounter. Patients undergoing routine outpatient procedures are at far less risk of serious injury from medical management than are those undergoing emergency surgery. Further, the more serious adverse events among outpatients, other than those that are immediately fatal, generally lead to hospitalization and thus are included in a pool of hospitalized patients. Finally, previous work has demonstrated that the yield of adverse events obtained from a sample of outpatient records is very low as compared with that from a sample of inpatient records.[1] One study of malpractice in emergency departments suggested a "bad event" rate of 0.0003 per emergency room visit.[2]

Patients suffering more serious complications of medical management in long-term care facilities are frequently referred to short-term, acute care hospitals. These cases were also included in our random sample. Long-term care per se, however, involves patients with chronic diseases in which the distinction between disabilities caused by the disease and by medical management are frequently difficult or impossible to recognize. Including these patients would have required special and costly methods beyond our resources. We regard this group of patients as deserving detailed study in the future.

A third practical issue was the validity of making a determination of injury and negligence on the basis of a pre-existing record. Compared with many records of physician office visits, the hospital record is usually more complete, compiled by different professionals, and required to meet minimum standards. The multiplicity of entries by various personnel provides

---

[1] See note 2 infra.

[2] Trautlein, Lambert & Miller, Malpractice in the Emergency Department -- Review of 200 Cases, 13 ANN MERG MED 709-11 (1984)

supplemental information that permits more accurate assessment of events. Laboratory test results and records of nursing care often provide primary or corroborative information of value.

Hospital records are far from perfect sources of information, however. Parts of the record may be missing. Important information may not be recorded. Adverse events may be deliberately described in a way that obscures what happened. Progress notes may describe a complication but not the assumed cause. Indeed, legal counsellors often recommend that physicians not state that an error occurred but that they merely describe the events. Notes may be illegible or confusing. Medical students, student nurses, and others with limited experience may describe an episode in such a way that it seems to be an adverse event when in reality it is not. Despite these shortcomings, hospital records are usually rich in detail, particularly for the more serious outcomes, and most major mishaps are likely to be recorded. The validity studies, described in detail in Chapter 5, provide evidence to support this view.

These practical considerations led us to define the stay in the acute care hospital as the basic unit of review. Because of the nature of adverse events, however, this does not limit our findings to events that occurred as a result of management during the hospitalization. A key premise of the Study is that medical management in the outpatient setting that causes serious injury to the patient will, in the large majority of cases, lead to hospitalization. Thus, enumerating adverse events that were caused by management outside the hospital but ultimately required hospital care, as well as those caused by management during hospitalization was assumed to capture the majority of serious adverse events.

Our universe of adverse events, then, consists of those occurring or discovered in the inpatient setting. For convenience, we chose our denominator to be total inpatient discharges, even though this does not correspond to the universe of encounters leading to the adverse events that we enumerate. Because a major focus of our Study is estimating total costs and compensation to the patients and litigation arising from adverse

events, this is not a serious limitation. Our overall adverse event rate is based on all medical management provided in hospitals and on management outside of hospitals causing adverse events that lead to hospitalization.[3]

A problem associated with defining our population was the temporal sequencing of medical management and discovery of the adverse events. The consequence of medical management might not be manifest and thus not discovered until after an initial clinical encounter has ended. For example, the adverse event might first be discovered during a subsequent hospital admission to a different institution. By counting in our random sample of hospitalizations those adverse events that occurred earlier and were first discovered in the sampled admission as well as those caused and discovered during the hospitalization, we compensated for those adverse events that occurred but were not discovered during the hospitalization.

Those adverse events that required multiple hospitalizations could be discovered and counted in more than one hospital admission. We resolved this potential problem of overcounting, faced by previous researchers, by adopting a set of counting rules that would produce an estimate of the incidence of adverse events. For details see Technical Appendix 5.III.1.

For the purpose of obtaining an incidence rate, we counted discoveries rather than occurrences.[4] However, the net number of discoveries should be equal to the occurrences of adverse events if (1) the underlying rate of events is stable over time during the period of our Study (approximately the decade of the 1980s)[5]; and (2) the loss to follow-up of patients who suffered an adverse event in New York and left the state for treatment is approximately matched by our sample patients whose adverse events

---

[3] A modest adjustment of overall rates would be needed to be descriptive of hospital practice apart from outpatient management. See Chapter 6 for a discussion.

[4] Discoveries refer to injuries first identified during the hospitalization, whether or not the cause of injury occurred during the hospitalization.

[5] We do not know how or whether the underlying rate changed. Although Chapter 1 shows claims rates increased, this may have been because a larger proportion of negligent actions resulted in claims.

occurred elsewhere but were discovered in New York hospitals. Because of the well known patient referral patterns into the state, the net
effect of this interstate movement would likely slightly overestimate the adverse event rate in New York.

2. Background: The Medical Insurance Feasibility Study[6]. Our study of hospital records was greatly helped by the only large earlier study of the subject, the Medical Insurance Feasibility Study, which was commissioned by the California Medical Association and the California Hospital Association in 1975. Not only did it serve as a model for our medical record study, but its director, Don Harper Mills, a lawyer-physician, was a consultant to our project in its early days.

Because Mills and his colleagues believed that the sensitive nature of their work would preclude their enlisting all hospitals that might be selected in a randomized sample, they picked 23 hospitals that would participate and that they believed were reasonably representative of the state as a whole. Their sample consisted of 20,864 records of inpatients who were discharged in 1974.

The Mills group defined a <u>disability</u> as a temporary or permanent impairment of physical or mental function or a medical event that caused economic loss even in the absence of impairment. A disability that had been caused by health care management was called a <u>potentially compensable event</u> (PCE). Within the study, causation was established when the disability was more probably than not attributable to health care management, which included acts both of commission and of omission. They identified three classes of PCEs: those resulting from treatment or procedures; those resulting from incomplete diagnosis or treatment; and those resulting from incomplete prevention. For such decisions, the study group considered the type and severity of injury, the circumstances in

[6]Medical Insurance Feasibility Study (D.H. Mills ed), San Francisco, CA (1977).

which it had occurred and was managed, and the condition of the medical records.

Although their study involved mainly hospitalized patients, Mills et al. also attempted to determine the frequency of PCEs in clinic records. In 928 such records they found only two PCEs, both of which were minor and produced temporary disability.

To avoid the expense of reviewing 20,000 charts in depth, the investigators defined 20 screening criteria that they found highly associated with PCEs. Using these criteria, trained personnel eliminated 50% of the sample charts as free of PCEs. The investigators themselves then reviewed the remaining 50%. Each record was presumably reviewed by only one investigator.

To classify PCEs, Mills et al. used a seven-grade severity scale. Minor disabilities that did not exceed 30 days in duration and did not require surgery were graded 3. The most severe disabilities, those that were fatal, were graded 7. The investigators decided whether a PCE was present, assigned a severity scale to it, and then made a judgment regarding tort liability.

In the sample of 20,864 records they found 970 PCEs, or an overall rate of 4.65%, including 3.81% from procedures, 0.69% from incomplete diagnosis or treatment, and the remaining 0.15% from incomplete prevention or protection. Of the PCEs 35.7% led to minor disability of 30 days or less and did not require surgery. Another 25.7% of the PCEs had minor disability that did not exceed 30 days but that did require surgery. Major disability, lasting for more than 30 days but less than two years, occurred in 18.6% of patients. Another 6.5% of PCEs resulted in minor but permanent partial disabilities: that is, conditions that were judged not functionally disabling. Major permanent partial disability occurred in 2.2%, and another 1.6% resulted in major or grave permanent disability. Death occurred in 9.7% of PCEs. Thus, almost 60% of PCEs caused only minor temporary disabilities, but nearly 10% of PCEs led to death.

Of the records reviewed, only 0.79% had evidence of negligence. But 42% of the PCEs that proved fatal were associated with negligence, whereas only 11.9% of the PCEs

leading to temporary disabilities were caused by negligence. Thus, negligence was found much more frequently in patients who suffered more severe PCEs.

Of the PCEs, 71.8% occurred in the operating room and another 12% in the patient's room. Specific procedures were responsible for 66% of PCEs, and drugs and other biologics for another 18.8%. Delayed diagnosis caused 2.7% of the PCEs, and misdiagnosis, another 2.2%. Post-procedural infections accounted for 29% of the PCEs. The largest single group of PCEs (6%) occurred in patients who underwent hysterectomies.

The California study was a major step forward in research on medical injury. As the first study of its kind and magnitude, it understandably had a few problems. For example, although several reliability tests were carried out on data retriever performance and on criteria adequacy, reliability and validity of the routine record reviews by physicians were not rigorously tested. (Reliability refers to the reproducibility of judgments of the reviewers; one measure of reliability comes from comparing the scores of multiple reviewers of the same record. Validity is an estimate of the truth in a judgment and is measured by comparing judgments made with two or more methods.) In addition, because the hospital and record samples were not randomly selected, the projections of total numbers of PCEs have been questioned.

The improvements in these areas that we were able to achieve were facilitated in important measure not only by lessons of the California experience, but with the consultative help of Dr. Mills. The other components of our Study -- the research concerning the economic effects of injury, the correlative studies of litigation records, the approaches to providers, and the interviews with physicians -- were not parts of the California research plan. Since publication of the California study, a few other studies of iatrogenic injury in the United

States have been carried out, but all are limited by their small sample sizes and their reliance on single reviews by clinicians.[7,8,9,10,11]

3. Technical Considerations. An early major challenge of the Study was to obtain a random sample of all hospitalizations in the year selected. (Recall that the California study was carried out in a volunteer sample.) As described in detail in Chapter 4, we used data from SPARCS to obtain a sample of 31,429 discharges from 51 distinct hospital facilities. The patients selected were representative of the state population in terms of race, payer class, age, sex and type of hospital. The latter were representative of urban and rural, teaching and non-teaching, for-profit and not-for-profit, and municipal and non-municipal institutions.

Obtaining participation of these 51 facilities was a major key to our success. In today's legal climate, health care providers are understandably concerned about any inquiries that might lead to law suits. Although the Department of Health and our Study group offered hospital administrators and their lawyers assurances of confidentiality, their responses contained a measure of skepticism. Our announcement that we planned to interview patients did not lessen such concerns. Nonetheless, 21 of the 51 hospitals originally selected responded positively almost at once to our invitation to participate. The others asked for additional reassurance concerning confidentiality and their own protection from legal liability. Two hesitated for almost a year, but ultimately all 51 hospitals originally

---

[7] Steel, Gertman, Crescenzi & Anderson, Iatrogenic Illness on a General Medical Service at a University Hospital, 304 N. Engl. J. Med. 638-42 (1981)

[8] Couch, Tilney, Rayner & Moore, The High Cost of Low-Frequency Events: The Anatomy and Economics of Surgical Mishaps, N. Engl. J. Med. 634-37 (1981)

[9] Lakshmanan, Hershey & Breslau, Hospitals Admissions Caused by Iatrogenic Disease, Arch. Inter. Med. 193-34 (1986)

[10] Abramson, Wald, Grennik, Robinson & Snyder, Adverse Occurrences in Intensive Care Units, 244 J.A.M.A 582-84 (1980)

[11] Brook & Stevenson, Effectiveness of Patient Care in an Emergency Room, 283 N. Eng. J. Med. 904-7 (1970)

selected in our random design agreed to participate. Their participation is likely testimony to the level of concern about malpractice issues on the part of New York providers, who appeared to agree with us that meaningful reform requires extensive and reliable information. In any event, their involvement was crucial to the success of our work, and we are most grateful to them.

Next was the issue of record retrieval. The frequency with which records move in and out of hospital record rooms means that finding large numbers of records poses a challenge in virtually every institution. Yet the record rooms of the 51 participating hospitals provided us at the outset with almost all (some with all) of the records we requested (overall, more than 96%). The nature of our Study, however, made it necessary to ensure as fully as possible that any records that could not be found were no different in profile from those that we reviewed. Therefore, we did a follow-up study four to six months later to retrieve information on the records not located at the time of our initial visits.

Of the records that could not be found on the first record room visit, about 50% were discovered and made available to us subsequently. We carried out full physician review of a random selection of these in order to estimate the adverse event rate among the missing records. This would have enabled us to make a realistic adjustment for bias due to missing records. As explained in detail in Chapter 6, no adjustment was needed, and we feel it is highly unlikely that our overall findings are misleading because of the missing records. We also insisted that hospitals supply <u>all</u> randomly selected records, including those segregated for quality assurance or risk management review.

In order to have adequate numbers of adverse events for the economic and other aspects of our Study, we oversampled records from Diagnosis Related Groups (DRGs) associated with certain high-risk specialties. Final estimates of adverse events and of

negligence for the entire population were adjusted to correct for this oversampling.[12]

Much effort was required to develop the record review process itself. The judgments asked of physician-reviewers were complicated and had to be completed in a reliable and valid fashion. Much of the preparatory work involved development of forms for guiding physician-reviews. Using these forms resulted in judgments that were generally both reliable and valid. (See Chapter 5.)

Because our review of hospitalizations was necessarily confined to medical records, the accuracy of our judgments depended on the accuracy and completeness of the record itself. We undertook several studies to evaluate the extent to which medical records fail to reveal adverse events. Our results indicate that in general medical records are reliable sources of information and that the majority of adverse events are detectable with our methods (Chapters 5 and 6).

Reviewers had to find in the medical record positive evidence of an adverse event before terming it such. Our Pilot Study and frequent spot testing throughout the main study on these issues demonstrated reliability and validity of a high order (Chapter 6). In our judgments about negligence we used as the standard what a competent physician would have considered reasonable in the setting involved.

C. Litigation Studies

Estimating the incidence of malpractice claims by patients in our Study presented administrative difficulties as well as technical challenges. As noted previously, our Study design precluded asking all patients directly whether they resorted to litigation. Further, we could not interview more than 10% of the patients in the Study within our time and budget limitations. Hence, we relied on external sources of data concerning malpractice litigation.

---

[12] See Chapter 4 for details.

Although the New York State Department of Health maintains a central repository for reporting claims, we needed to ascertain whether the New York data base was complete and then to repair any deficiencies. We met first with officials from the Office of Professional Medical Conduct (OPMC) to determine the shortcomings in claims reporting and to plan a timetable for data collection. Next, we met twice with representatives of the Insurance Department to request assistance in encouraging compliance with claims reporting statutes and regulations. Finally, we made independent overtures to several hospitals to gather claims data that had not been previously reported to Albany. This process was begun in December 1987 and completed in August 1989.

To determine the frequency of litigation, we developed methods to link our sampled medical records with records in the malpractice claims data base. With computerized hospital discharge data and the more recent malpractice claims data, we employed computer-assisted exact matching. For the 30,000 older claims and for those hospitals with poor claims reporting records, however, we had to rely on manual matching of hospital discharges and claims abstracts. In addition, when computer linkage of records indicated matches of discharges and claims, we verified the matches via hardcopy checks. Linking clinical and legal data was a major undertaking for the Study's support staff and benefited from enormous cooperation among the Health Department, the Insurance Department, insurers, and hospitals. As a result, the Study was able to avail itself of an extraordinarily extensive data base on malpractice litigation. Chapter 7 describes these data and their implications in detail.

D. Patient Interviews

The interview survey is, to our knowledge, the first attempt to collect data on post-hospitalization costs and compensation to victims of medical injuries. We described in the prior section the efforts we made to locate and review hospital records for our sample in order to estimate the incidence of adverse events caused by medical intervention in 1984. In order to document the economic consequences of these injuries to the patients affected,

we had to locate and interview the patients five years later. Since some of the patients had died, many had moved, and all had to be persuaded to participate in an hour-long interview, the patient survey posed logistical challenges in its execution.

The patient interview sample was drawn from the patients identified in the record survey. There were 3,341 unique discharges in the patient interview sample. The sample consisted of a group of patients who suffered adverse events during medical management and another group of patients who were not injured. The uninjured patients were matched to the adverse event patients on the basis of diagnosis, age, sex, race, hospital, Medicaid and working status.

The survey instrument contains questions concerning seven time intervals: the six months prior to the patient's hospitalization; the period of hospitalization during 1984; the interval from the date of discharge to the end of calendar year 1984; each of the years 1985, 1986, and 1987, and the first six months of 1988. The survey instrument, which is reproduced in Appendix 8A of this report, consists of ten sections. Every respondent was asked questions on medical care utilization and health insurance (sections 1 and 2) and on the amounts and sources of non-wage income (section 10). The choice among sections 3 through 9 is determined by the patient's primary usual activities (that is, work for wages, homemaking, unemployed/disabled, retired, student, preschool child) during the six months preceding admission to the hospital in 1983-1984.[13]

We began to design the survey instrument in 1986. In 1987, during the Pilot Study, we tested it on 295 patients who were discharged from New York hospitals in 1983. That Study revealed serious problems in the process of locating potential respondents and some design errors in the survey instrument. Of the 293 patients selected for the Pilot Study interview, 186 (64.5%) could not be found. Although 85.7% of the persons who were located agreed to be interviewed, the very low locate rates

[13] Because some of our patients who were discharged in 1984 had been admitted in 1983.

produced a final interview rate of 58.0%, which is below the norm accepted for survey research.

We reviewed the non-located cases and compared them to the medical records from which the locational information had been collected. The procedures for conducting the medical record reviews were changed to eliminate the problems that we discovered. We emphasized the importance of locating patients in the training of medical record reviewers and made unannounced visits to medical record review teams in the field to sample the data being collected.

Information to help us find patients was also obtained from Hanes Reverse Directory Service, New York State Medicaid files, the New York State Department of Motor Vehicles, and commercial credit reporting organizations. Former employers and neighbors were contacted, and New York State provided us with letters of authorization for obtaining location information from superintendents of public housing projects.

These efforts permitted us to locate 2,968 (89%) of the sample patients in the larger Study. Of the 2,968 persons who were located, 37 were not eligible for the interview because they had moved to a foreign country (N=14) or because they and their next-of-kin had died (N=23). An additional 59 persons had moved outside of the state. Thus, the sample of eligible, located cases numbered 2,872 persons.

The next objective of the survey process was to convince people to be interviewed. Our first concern was to eliminate barriers to interviews that might be correlated with the respondents' socioeconomic status. Face-to-face field interviews were scheduled for many persons who could not be interviewed by telephone. A Spanish language version of the survey instrument was produced.

One of our problems was that the study design precluded our revealing to the interviewees that the data were to be used to study, among other things, medical malpractice. Neither the interviewers nor the potential respondents were aware that such a study was among our objectives. Subjects could only be told that we were studying "the consequences of hospitalization." In the

early weeks of the survey, refusal rates exceeded 20%. Interviewers were given additional training on dealing with refusals and reluctant subjects were offered a small payment for their participation. Many of the initial refusals were converted to interviews, and the refusal rates of subsequent respondents were far lower. The overall refusal rate among located patients was 10.4%. If non-respondents are significantly different than respondents in the characteristics that were used to select the sample, the survey results will be biased accordingly. In what follows we compare the characteristics of the respondents and non-respondents to evaluate whether any of our results could be biased.

Eighty-six field interviews were completed. A total of 2,326 interviews (field and phone) were completed for patients who were located and eligible, yielding a response rate (total interviews/total located and eligible = 2,326/2,872 patients) of 81%. The total interview rate was 70% (total interviews/total eligible = 2,326/3,341).

The interview data were collected by Mathematica Policy Research, Inc. of Princeton, New Jersey. The Pilot Study interviews were conducted by the National Opinion Research Center of the University of Chicago.

The interview data were merged with hospital record information on the patients' 1984 hospitalizations and have been matched with the complete Health and Insurance Department reports of tort claims for the years 1983 through July 1989. The combined data sets serve as the basis for our estimates of the economic consequences of iatrogenic injuries.

The interview survey provides a unique description of the economic consequences of iatrogenic injuries. It includes data that can be used to determine the extent to which the costs of malpractice are paid by social insurance programs, such as Medicare and Medicaid, or by private health and disability insurance. The data can also serve as the basis for estimates of the costs of proposed alternatives to the current tort liability approach.

E. Studies of Deterrence

Considerable effort was directed at determining to what extent the tort system serves to deter physicians and other providers from negligent behavior. For this purpose, we surveyed physicians and carried out an econometric study of deterrence.

1. Surveys of Physicians. The surveys were designed to seek information about the messages sent by the tort system in order to gain insights into its role in guiding physician behavior and in preventing patient injuries. For this purpose we approached physicians in two ways: a mailed survey and in-depth personal interviews.

The disadvantages of the personal interviews include the small sample size (they are time consuming and costly, thereby reducing the number that we could do) and difficulty in generalizing from the data. We decided to undertake the interviews despite these drawbacks, because they offer an opportunity to obtain extremely detailed data on physicians' perceptions of the tort system and of their quality assurance activities. We treated the interviews as case studies.

The mailed survey of doctors allowed us to reach a wider pool of physicians and to focus on a few key issues. Data from the mailed survey could be analyzed quantitatively and subjected to standard statistical testing techniques.

Both survey formats presented special challenges in sample selection. To provide a wide range of opinions and perceptions, we focused on three broad groups: specialties at high risk, average risk, and low risk of a malpractice suit. We also stratified each sample by geographic area to capture the effect of location on perceptions and attitudes. Further, for the mailed survey, we stratified by claims history: that is, whether the doctors had claims against them reported to the State Health Department. Ultimately, we received 739 mailed interviews and conducted 47 face-to-face interviews.

Both surveys present a limited view of incentives under the tort system insofar as they are from physicians, but we accepted

the limitation because of the importance of physicians' perceptions. Indeed, the perception of incentives largely shapes the behavior that ultimately affects patient care.

2. Econometric Study. An ideal study of deterrence induced by the tort system would compare the frequency of negligent adverse events in similar jurisdictions, some of which had the tort system and some of which did not. Moreover, it would be randomly determined which jurisdictions had the tort system; or at least the presence of the tort system would not be determined by the frequency of negligent events.

Clearly, such a study cannot be done because all jurisdictions are subject to the tort system. Thus, our study could not observe the amount of negligence that would occur if the tort system were replaced with something else (for example, a no-fault system).

In the absence of such a study, can anything be done that will suggest the amount of deterrence induced by the tort system? We believe the best that can be done is to study the consequences of the existing variation in the threat of a lawsuit; hospitals and medical staff vary in the likelihood that a lawsuit will be brought, and there may well be some behavioral consequences for both injury rates and cost. In particular, do areas where the threat of a suit is relatively high exhibit lower rates of negligence, lower rates of injury, and higher cost?

In assessing the amount of tort threat, we face a potential problem of dual causality. A high rate of negligence may induce a high threat of suit; on the other hand, a high threat of suit may induce a low rate of negligence. We discuss this problem further in Chapter 10.

IV. Concluding Note

We have described several problems that we overcame in the shaping and execution of our research project. The methods we used and the results we achieved are presented in detail in the chapters that follow. The work will also be described in

articles that will be submitted to a range of peer-reviewed scholarly journals.

Chapter 4

THE NATURE AND INCIDENCE OF MEDICAL INJURY: SAMPLING FOR THE RECORD REVIEW

Summary

In order to make our results generalizable to the entire population of hospital discharges in New York, we drew a probability sample of 31,000 hospital records. Our ability to obtain such a sample was made possible by the availability of the SPARCS data system. The basic sampling design of the Study is an implicitly stratified, systematic, two-stage cluster sample of discharges. We first selected hospitals with probabilities proportional to the number of non-psychiatric discharges; records within hospitals were then selected with three different sampling frequencies determined by patient age and diagnosis-related group (DRG). Using SPARCS information on patient discharges, we drew a sample with a distribution that conforms closely to the population on important hospital and patient characteristics.

I. General Sampling Considerations

One approach to our Study could have been to examine all 2.7 million medical records from 1984 for evidence of adverse events, including negligent adverse events, but this would have been enormously expensive. Instead, we designed a sampling strategy to allow us to estimate what the results would have been if we had examined all records and interviewed all living patients. Our sample comprised slightly more than 1% of the hospitalized patients. This section describes the purposes and techniques of the sampling and provides summary data for comparisons between the sample and the population.

Many of the assumptions that guided our development of the sample are based on two other studies of adverse event rates: the California Medical Insurance Feasibility Study[1] and our own Pilot Study, completed in three New York hospitals in 1986.[2] For the Pilot Study, we surveyed 2,800 admissions selected to contain an increased proportion of high-risk cases. Because a primary objective of the current Study was to estimate the costs associated with adverse events, the total number of sampled discharges was predicated on the volume of adverse events yielded rather than a desired level of precision for the incidence rates of adverse events. A goal of 1,500 adverse events was set as a target, based in part on available wage loss data and in part on information from the Pilot Study about the distribution of economic losses in the population.

Information about the likely rate of adverse events was available from both the California study and Pilot Study. The California study indicated an approximate figure of 5% for "potentially compensable events," implying a need to sample 30,000 records to achieve a total of 1,500 adverse events. Our Pilot Study yielded a slightly higher percentage, as we might expect, because it deliberately sampled a high-risk population. We chose not to rely on the possibility of a higher adverse event

---

[1] Report on the Medical Insurance Feasibility Study, San Francisco: California Medical Assn. (D.H. Mills ed. 1977)

[2] Medical Practice Study Group. Boston: Medical Practice Study Group, 1987.

rate because of our concern that the non-response rate for the patient interview survey would be high. In the Pilot Study, the response rate in the patient survey was only 58%. A goal for the main Study was set at 70%. These considerations led us to plan for a review of about 30,000 records in order to yield the goal of 1,500 adverse events.

The second issue was the necessity of reviewing charts on site in active medical record rooms. This dictated a two-stage cluster design: we first randomly selected hospitals (or clusters) and then randomly selected records within each chosen hospital. Although the efficiency of a cluster design increases with increasing numbers of clusters, the administrative difficulties, time delays and costs of dealing with each hospital limited the feasible sample size to 30 hospitals. By using a linkage process to form some hospital clusters, we were able to sample 51 different facilities.

A third issue was the preference for a self-weighting sample because the data would be subjected to numerous secondary analyses. In a self-weighting design, each observation in the sample represents the same number of discharges, and raw rates and ratios calculated for the sample apply to the population. A competing preference was that of maximizing the yield of adverse events. The variation in adverse event rates across individual hospitals was not known, but both the California study and the Pilot Study gave rough estimates of adverse event rates in major diagnostic categories and patient age groups. Thus, the final sample was chosen to be self-weighting with respect to hospitals but not with respect to patient age or diagnosis. This choice complicated analysis only slightly because modern computer software handled the varied subgroup weighting within each hospital or cluster easily. Thus, we were able to generalize without difficulty.

The survey design was facilitated by the fact that New York State maintains the Statewide Planning and Research Cooperative System (SPARCS), a computerized file of patient discharges from

all hospitals in the state.[3] This file provided the basic sampling frame for our survey as well as access to detailed population characteristics. SPARCS, a division of the Health Department, receives and processes inpatient data from all hospitals in New York State. The data set for a calendar year includes all patients discharged in that year even if they were admitted in a prior calendar year and excludes data for patients entering a hospital in that year but discharged in subsequent years.

SPARCS collects data from two sources: the Discharge Data Abstract (DDA) and the Uniform Billing Form (UBF). The DDA data, based on the Uniform Hospital Discharge Data Set, come from the hospitals' medical record departments and are validated and carefully screened for duplicate submissions. The UBF data represent data gained from the uniform claim form used by all third-party payers in New York State. After a patient discharge, the hospital billing department distributes copies to the appropriate third-party payers. SPARCS gathers UBF data only after Medicare, Medicaid, or other insurers have processed the claims, thereby allowing SPARCS to benefit from existing checks and edits.

DDA and UBF data sets have common elements that allow matching and merging of the two files. SPARCS estimates that approximately 15% of DDA and UBF data remain unmatched. In all cases we chose to use DDA records to count discharges because DDA records are generated shortly after discharge and because a UBF submission does not always represent an inpatient discharge. Whenever possible, we improved upon the SPARCS matching/merging process by examining key fields and correcting obvious keypunching errors.

Three other considerations led to the final list of hospitals and medical records to be sampled (the sampling frame). First, only hospital discharges in 1984 were included. Because our study design called for a review of malpractice claims

[3] New York State Department of Health. Statewide Planning and Research Cooperative System: Annual Report Series, 1984, v. 1. Albany, N.Y.: DOH, 1984

arising out of adverse events, the time from medical management to review had to allow for the usual delay in filing malpractice claims. We were also sensitive to the perceived risk that our study would induce patients to litigate. Therefore, we wanted a study period for which the statute of limitations for filing suit had run out.

Second, because of the difficulty of distinguishing between treatment effects and underlying disease as the cause for an undesirable outcome, our Study was deliberately limited to acute, short-term hospital care excluding psychiatric and long-term care. Finally, hospitals operating in 1984 but closed prior to the beginning of our Study were excluded because obtaining the necessary cooperation from these hospitals would be impossible.

The process of medical record selection proceeded in several stages. First, we determined from Department of Health and American Hospital Association data the list of all acute-care hospitals in New York State still operating in 1988. Using this sampling frame, we next stratified this list according to hospital ownership, teaching status, and location. Then beginning with a random starting point generated by computer, we selected 31 hospital clusters systematically. Stratification and systematic sampling ensured that random selection would produce a sample of hospitals representative of all types. Finally, again using a computer-generated random number, we chose individual discharge records systematically from these facilities after stratifying by age and diagnostic group. As part of this sampling process, each discharge received a sampling weight that signified the number of discharges represented by this sample. Because of random selection techniques and the application of sampling weights in our analyses, the Study could extrapolate findings based on the sample to the entire state population of acute-care patients.

Sections II and III describe in more detail the sampling techniques. Section IV compares the characteristics of the sample actually drawn with those of the population of all patients in acute-care hospitals in 1984.

II. Selection of Hospitals -- The Sampling Frame

Our desire to have information on the full spectrum of New York hospitals, including small rural hospitals, and the concern that adverse event rates may vary according to hospital characteristics led us to stratify hospitals by geographic location, ownership and teaching status. In arriving at the sampling frame for hospitals, we first classified hospitals according to the stratifying variables and acute-care status. To avoid distortion due to specialty hospitals or unique facilities, some institutions were combined into one cluster before drawing a sample.

Guided by these parameters, we reviewed a list of acute-care facilities in the New York Department of Health Facilities Directory, July 1984. Using the American Hospital Association's Guide to Health Care Institutions, 1985 edition[4] (listing 1984 data), as well as length of stay data from the Department of Health, we identified some of these institutions as being predominantly long-term facilities or psychiatric hospitals and deleted them from the data base. The institutions that had closed during or after 1984 were also deleted. Hospitals that changed names or merged with other facilities were maintained in the frame with their new names and status. After these modifications, 270 institutions remained.

To identify administrative control, the Department of Health indicated whether institutions with separate facility indicators were administratively joined as a single entity. The rural designation was based on whether a hospital was located outside a Metropolitan Statistical Area (MSA) as defined by the U.S. Department of Commerce, Bureau of the Census.

An analysis of medical school affiliation and approved residency positions led us to three levels of hospital teaching status: university teaching, affiliate teaching, and non-teaching. As university teaching hospitals we chose the 13 facilities designated by the state's medical schools as their

[4] American Hospital Association, Guide to Health Care Institutions. Chicago: AHA, 1985.

primary clinical centers. All remaining hospitals with at least five approved residency programs and five specialty hospitals with large numbers of residents on the staff were classified as affiliate teaching. For sampling purposes, we grouped all teaching hospitals together. As we outline in Chapter 6, for analysis of data by hospitals we used the university, affiliate, non-teaching classification. (See Technical Appendix 4.II.1.)

Each specialty hospital was joined with a general care facility to form one cluster. After considering academic affiliation and joint residency programs, we looked for physical links such as common buildings or walkways. Finally, if no administrative or physical links were present, we chose a nearby large general hospital. One maternity facility was not clustered because it is a proprietary facility with no neighboring proprietary facilities. By grouping specialty hospitals to form heterogeneous clusters, we increased the potential efficiency of the two-stage sample design.[5]

The joining of special to general hospitals reduced the sampling frame to 240 principal hospitals (or clusters) consisting of 270 facilities. For purposes of stratification, the geographic, teaching, and ownership characteristics of the principal hospital in a cluster defined the status for all hospitals in that cluster. Once the 240 hospital clusters were determined, we organized them into strata determined by location, teaching status, and ownership. According to the method of implicit stratification, hospitals were listed in an order dictated by stratifying variables and then sampled systematically. To create the list, we first identified the county location of the principal hospitals and determined whether the county belonged to an MSA. One major teaching hospital in a rural area was joined with the nearest MSA group because it was not representative of rural (non-MSA) institutions.

We then grouped all principal hospitals into five strata: (1) MSA private, nonprofit hospitals; (2) MSA proprietary

[5] Kish, L. Survey Sampling 161-64 (1985).

hospitals (there were no rural proprietary institutions); (3) all rural (non-MSA) hospitals regardless of ownership; (4) government-owned MSA institutions; and (5) N.Y. Health and Hospitals Corporation (NYHHC) institutions. Within each of these strata, we further ordered hospitals by teaching status (using only teaching/non-teaching designation), listing the non-teaching institutions first. Since no rural hospitals or proprietary hospitals were teaching hospitals and all NYHHC facilities were teaching institutions, this made a total of seven strata. The number of principal hospitals, actual facilities, and discharges in each of these seven major strata are shown in Table 4.1.

Table 4.1
Stratification of Acute-Care Facilities 1984

Strata Characteristics		Number of Hospital Clusters	Number of Facilitities	Number of Non-psychiatric Discharges
1	MSA, private nonprofit, teaching	46	69	1,136,680
2	MSA, private nonprofit, non-teaching	100	105	846,877
3	MSA, proprietary, non-teaching	18	18	123,783
4	Non-MSA hospitals	53	55	239,229
5	MSA, government ownership, teaching	6	6	79,176
6	MSA, government ownership, non-teaching	6	6	22,040
7	NYHHC	11	11	224,078
	TOTALS	240	270	2,671,863

Within the seven strata, we ordered the hospitals by geographic county. Beginning with Suffolk County on Long Island and winding first west, then north to the Canadian border, then

south again and ultimately to Erie, we ordered the counties. The purpose was to ensure a balanced geographic distribution of hospitals as well as some geographic proximity for institutions next to each other on the list. At the end of the non-teaching substratum on the list, we reversed the direction of the Suffolk to Erie ordering of counties so that the last non-teaching hospital would be geographically near the next hospital on the list, a teaching facility. The number of non-psychiatric discharges for each institution was obtained from the SPARCS 1984 data as of March 1987. A computer program then aggregated the net number of discharges for each institution in a cluster and assigned that discharge total to the principal hospital. The net total of discharges was 2,671,863. With a goal of conducting 30,000 medical record reviews, we selected 1 in 86 discharges, giving a target sample size of 31,068, to compensate for an anticipated non-locate rate of approximately 3%.

The basic design called for a self-weighting probability sample drawn in two stages. We first sampled 31 hospital clusters from the population of 240 defined clusters, then sampled approximately 1,000 discharges from each of the 31 clusters. To obtain a self-weighting sample, we chose clusters with probability proportional to their total number of discharges. About 80 of the 240 clusters had fewer than 5,000 discharges and some had fewer than 1,000 non-psychiatric discharges in 1984. These facts, plus the desire to ensure adequate representation of small hospitals and increasing the number of hospitals sampled, led us to refine the basic sampling technique to accommodate undersized units.[6] Any hospital with fewer than 8,000 discharges was linked (for sampling purposes) with adjacent hospitals until the total discharges in the group reached at least 8,000. The figure 8,000 was chosen, after some simulations, to target the expected yield of small (<5,000 discharges) hospitals at about 10, as opposed to about 3 expected if no grouping by size were made.

[6] Kish, L. , note 5 supra, 244

In order to maintain probability-proportional-to-size sampling, the undersized hospitals should be grouped with adjacent units prior to drawing the samples. We did the linkage after the hospitals were selected. The net result was a slightly lower probability of selection for hospitals with fewer than 8,000 discharges. We corrected by computing the actual probability of selection for all hospitals and reweighting accordingly. (See Technical Appendix 4.II.2.)

The final 31 hospital clusters were selected by systematic sampling of the ordered list of 240 principal hospitals. Beginning with the first hospital cluster on the ordered list described above, each cluster was assigned a cumulative measure of size equal to its 1984 discharges plus the cumulative total discharges (CTD) for all previous clusters on the list. Then, to prevent accumulation of rounding errors, each institution was assigned a number of sampling units (SU), where for the jth hospital,

$$SUj = (CTDj - CTDj\text{-}1) / 1{,}000$$

Any fractional SUs were rounded to the nearest integer. We then calculated for each hospital the cumulative sampling units (CSU), defined as the sum of all sampling units for it and any hospitals above it on the list. To start the sampling process, we selected a random number, x, between 1 and 86. The first hospital selected was defined as the first principal hospital on the list with $x \leq CSU$. Thus, at least one hospital would be selected from the first 86,000 (approximate) discharges on the list. The remaining hospitals were selected systematically by adding 86 to x to determine the next CSU cutpoint (i.e., the second hospital selected was the first on the list to satisfy $x + 86 \leq CSU$, etc.[7]). This procedure identified 31 hospital clusters comprising 51 separate facilities.

[7] Kish, L. Survey Sampling (1965), 234-237.

III. Selection of Hospital Records

In a two-stage, probability-proportional-to-size design, the number of sampling units for a hospital or cluster determines the sampling fraction within the cluster. The exact number of cases sampled in each cluster was obtained by dividing the total number of non-psychiatric discharges for the cluster by the total number of sampling units for that hospital or group. Thus the basic sampling rate for a cluster was 1/(sampling unit), or 1/SU. If a cluster actually consisted of several different facilities, the number of discharges sampled from each hospital in the cluster was proportional to the number of discharges they contributed to the cluster.

Although a self-weighting sample is desirable to simplify analysis, our desire to maximize the number of adverse events to study economic consequences and deterrence led us to stratify medical records within hospitals and to use disproportionate sampling within these strata. We chose the strata as follows. First, although the elderly tend to produce a disproportionate number of adverse events (they accounted for 20% of the discharges but 36% of the adverse events in our Pilot Study), the variation in economic losses is less among the elderly, mainly because work loss is a relatively minor consequence of injury for most elderly. More important for the cost study is an adequate number of workers and children who suffer the adverse events, because of the large economic consequences due to disability and/or loss of earning potential in these groups. This led us to sample the elderly at 50% of the base rate, following the Medical Insurance Feasibility Study.[8] The elderly group included all patients age 70 and over, regardless of diagnosis. The lower sampling rate for the elderly increased the numbers of younger patients in the sample since the overall total (31,000) remained fixed.

Second, our Study goals necessitated special attention to certain subspecialties. The deterrence study was deliberately

[8] See D.H. Mills, note 1 supra

designed to involve six specialties thought to be at high risk for adverse events. The six high-risk specialties are neurosurgery, orthopedic surgery, thoracic surgery, peripheral vascular surgery, cardiac surgery, and urology. Data from the Pilot Study and the Medical Insurance Feasibility Study suggested an average adverse event rate as high as 10% might be expected for these specialties. Because we lacked direct information on physician specialties at the time we sampled records, we identified selected DRGs associated with those specialties and oversampled them. To obtain a sufficient number of records per hospital for the deterrence study, we oversampled the 6 specialties by a factor of 3. The factor 3 was chosen to make the average expected number of records per hospital cluster in the deterrence study at least 8 for the rarest groups (thoracic, peripheral vascular and cardiac surgery). The average expected number of records per hospital was 17 for neurosurgery and considerably higher for the other specialties.

Obstetrics and neonatology required special attention as well. Current data suggest that malpractice claim rates per physician are high for these specialties although overall adverse event rates are low due to the vast number of normal deliveries. Without weighting, the number of obstetrics and neonatology records would be sufficient for the deterrence study, but the expected number of adverse events in many hospitals would be near zero. As a result, we divided obstetrics and neonatology further by DRG into two categories: (1) low risk-obstetrics and neonatology, which includes all normal deliveries and newborns (DRGs 371, 373, 374, 379, 380, 381, 382 with a total of 299,758 discharges in 1984) and (2) high-risk obstetrics and neonatology (DRGs 370, 372, 375, 376, 378, 383, 384, with 57,548 discharges in 1984). The high-risk obstetrics and neonatology and DRGs were also oversampled by a factor of 3. Low-risk obstetric/neonatology cases were sampled at the elderly rate of 50%.

In summary, the final within-hospital design required coding of records into three strata: one sampled at 3 times the base rate; one sampled at one-half the base rate; and the third at the

base rate. The overall base rate of 1/86 was then adjusted slightly downward to maintain the desired number of approximately 31,000 records (see Technical Appendix 4.II.2).

Records within hospitals were selected systematically using a different random start for each of the three strata, the sampling interval being a function of the stratum and hospital SU. Using this method, a total of 31,429 discharges were selected for review. Each discharge selected for the sample carried a sampling weight equal to the number of acute patients in the total population represented by this sampled discharge. The weight of a sampled record equals the inverse of its probability of selection. In our two-stage sampling design, the probability of selecting a specific discharge depended on the probability of selection of the hospital and the probability of selection of a given record within the hospital. The computer program that systematically selected records within hospitals also assigned a weight to each record. Details of computation appear in Technical Appendix 4.II.2.

Weights for specific records within the sample (unadjusted for missing records) vary from 22.8 to 1009.1. The mean weight is 85.9, the median is 88.3, and 80% of the weights fall in the range of 31 to 189. A sampled record with a weight of 86, for example, represents 86 discharges from our population of 1984 discharges. When the weights for all sampled records are totalled, they equal approximately 2.7 million, the number of acute hospital discharges in New York State in 1984. Using sample weights in our calculations and analyses, we can make inferences about the population.

IV. Comparison of Sample and Population Characteristics

In this section we compare the sample distribution with the population distribution on a number of key hospital and patient characteristics. Tables 4.2 - 4.5 give the hospital characteristics (size, ownership, location and teaching status) while Tables 4.6 - 4.10 compare the age, race, sex, MDC and insurance class characteristics. Because the only substantial minority groups were blacks and Hispanics, other minority groups

(primarily Asians) were combined with whites. We compare the weighted sample proportions with the corresponding population proportions.

Table 4.2
Discharges by Facility Size

	SAMPLE		POPULATION	
FACILITY SIZE	Number of Facilities	Percentage of Discharges (weighted)	Number of Facilities	Percentage of Discharges
Under 8,000	24	19.4	139	22.3
10,000 to 19,999	16	50.4	105	51.5
20,000+	11	30.2	26	26.2
TOTALS	51	100.0	270	100.0

Table 4.3
Discharges by Ownership

	SAMPLE		POPULATION	
HOSPITAL OWNERSHIP	Number of Facilities	Percentage of Discharges (weighted)	Number of Facilities	Percentage of Discharges
Non-profit	40	78.6	220	81.9
Investor-owned	3	6.9	18	4.6
Government, non-federal	8	14.6	32	13.5
TOTALS	51	100.0	270	100.0

Table 4.4
Discharges by Geographical Area

	SAMPLE		POPULATION	
GEOGRAPHICAL AREA	Number of Facilities	Percentage of Discharges (weighted)	Number of Facilities	Percentage of Discharges
Upstate Urban	18	34.4	95	30.6
Upstate Rural	4	6.7	55	9.0
Nassau-Suffolk-Westchester	8	21.0	41	17.5
New York City	21	37.9	79	42.9
TOTALS	51	100.0	270	100.0

Table 4.5
Discharges by Teaching Status

TEACHING STATUS[9]	SAMPLE		POPULATION	
	Number of Facilities	Percentage of Discharges (weighted)	Number of Facilities	Percentage of Discharges
Non-teaching	25	50.1	186	46.7
Affiliate Teaching	22	38.0	71	42.3
University Teaching	4	11.9	13	11.0
TOTALS	51	100.0	270	100.0

Table 4.6
Discharges by Patient Age

PATIENT AGE	SAMPLE		POPULATION	
	Number of Discharges (unweighted)	Percentage of Discharges (weighted)	Number of Discharges	Percentage of Discharges
Newborn	3,699	9.5	244,374	9.1
0-15	3,190	8.4	247,492	9.3
16-44	11,545	34.2	931,103	34.8
45-69	10,122	26.8	711,225	26.6
70+	2,873	21.1	537,669	20.1
TOTALS	31,429	100.0	2,671,863	100.0

[9] For a description of our definition of teaching status, see Technical Appendix 4.II.1.

Table 4.7
Discharges by Patient Race

PATIENT RACE	SAMPLE		POPULATION	
	Number of Discharges (unweighted)	Percentage of Discharges (weighted)	Number of Discharges	Percentage of Discharges
Black	5,218	15.5	390,321	14.6
Hispanic	1,834	5.3	211,152	7.9
White/Others	24,377	79.2	2,070,390	77.5
TOTALS	31,429	100.0	2,671,863	100.0

Table 4.8
Discharges by Patient Sex

PATIENT SEX	SAMPLE		POPULATION	
	Number of Discharges (unweighted)	Percentage of Discharges (weighted)	Number of Discharges	Percentage of Discharges
Female	17,135	59.6	1,583,318	59.3
Male	14,294	40.4	1,088,545	40.7
TOTALS	31,429	100.0	2,671,863	100.0

Table 4.9
Discharges by Patient Diagnosis

PATIENT DIAGNOSIS	SAMPLE		POPULATION	
	Number of Discharges (unweighted)	Percentage of Discharges (weighted)	Number of Discharges	Percentage of Discharges
Cardiology	2,252	10.3	266,405	10.0
Dentistry/Oral Surgery	191	0.7	18,993	0.7
Dermatology	90	0.4	10,152	0.4
Endocrinology	424	1.9	45,567	1.7
Gastroenterology	1,354	5.8	159,739	6.0
General Surgery	2,933	11.8	302,322	11.3
Gynecology	1,564	5.7	145,968	5.5
Hematology	201	0.8	23,247	0.9
Neonatology (incl newborns)	3,888	9.7	250,460	9.4
Nephrology	400	1.7	45,302	1.7
Neurology	981	4.4	109,848	4.1
Neurosurgery	756	1.1	30,335	1.1
Obstetrics	3,272	12.6	355,780	13.3
Oncology	918	3.8	94,354	3.5
Ophthalmology	560	2.6	79,146	3.0
Orthopedic Surgery	4,288	6.2	171,609	6.4
Otolaryngology	897	3.3	100,165	3.7
Pulmonary Disease	1,348	5.9	158,209	5.9
Rheumatology	132	0.6	13,999	0.5
Cardiac/Thoracic Surgery	885	1.2	30,794	1.2
Urology	2,384	3.7	97,321	3.6
Vascular Surgery	476	0.8	21,064	0.8
General Medicine (and other diagnoses)	1,235	5.0	141,084	5.3
TOTALS	31,429	100.0	2,671,863	100.0

Table 4.10
Discharges by Patient Insurance Class

PATIENT INSURANCE CLASS	SAMPLE		POPULATION	
	Number of Discharges (unweighted)	Percentage of Discharges (weighted)	Number of Discharges	Percentage of Discharges
Self-pay	2,408	7.0	190,633	7.1
Medicare	5,973	28.9	750,497	28.1
Medicaid	4,604	13.6	410,589	15.4
Blue Cross (and others)	18,444	50.5	1,320,144	49.4
TOTALS	31,429	100.0	2,671,863	100.0

As expected, there is good agreement between sample and population distributions. The agreement is considerably better for patient than for hospital characteristics, however. This is explained by the two-stage design. Although the number of records is large, the number of primary sampling units (31 hospital clusters) is relatively small. For only two patient characteristics does the sample proportion deviate from the population proportion by more than 1%: the proportion of Hispanic discharges (Table 7) and the proportion of Medicaid discharges (Table 10). This occurs because the proportions of both of these characteristics (particularly Hispanic) vary considerably from hospital to hospital. Thus the two-stage sample design is relatively less efficient for estimating these proportions.

TEACHING STATUS OF HOSPITALS IN NEW YORK STATE

We needed a formal definition of hospital teaching status for purposes of sampling and data analysis by hospital characteristics. Teaching status is not an immutable, widely recognized characteristic. To ensure consistency in our results, we chose to adopt a three-level definition of teaching status: university teaching, affiliate teaching, and non-teaching. This appendix summarizes the basis for the chosen classification and the data we collected on teaching status.

I. Data sources

Available data contained ambiguities and missing values. We found no consensus among the data sources as to definition or criteria for determining teaching status. In addition, the hospitals were not static: some facilities changed names and affiliations over time. In arriving at our classification, we used the following sources.

SEDI (New York Department of Health). The 1984 SEDI index,[1] supplied by the Office of Health Systems Management, Bureau of Health Economics, classifies a teaching hospital as any facility with at least 5 approved residency programs (e.g., surgery, internal medicine, obstetrics).

Hospitals designated by medical schools as their primary clinical facility. We also received a list of hospitals by peer groups from the Department of Health, formed by a combination of teaching status, ownership, and location. One of these groups

[1] J. Herbst, Assistant Director, Bureau of Health Economics. Letter to Medical Practice Study, May 11, 1987.

includes hospitals designated by the medical school as their primary teaching hospitals. Table 1 flags those 13 facilities. We have termed them the university teaching hospitals.

American Hospital Association Data. From the 1985 American Hospital Association Guide (listing hospital status as of 1984), we obtained separate indicators of whether (1) a hospital had approval to participate in residency training by the Accreditation Council for Graduate Medical Education (Jan. 1985) (designated AHA3 in Table 1), and (2) a hospital had a medical school affiliation reported to the American Medical Association (Jan 1985) (designated AHA5 in Table 1).

Approved Residency Positions (DOH). In February 1987, the Department of Health provided us with a list of all approved residency positions and the number of full-time equivalents (FTE) by hospital. These data carried the following caveat. First, figures were for early 1987, and therefore could differ from the totals for 1984. Second, DOH attached the following caution:

> While hospitals reported 12,933 approved and non-approved positions to us, these same hospitals can only account for 12,617 FTEs. The 12,933 figure is reliable. However, through numerous conversations with many hospitals concerning their rotation schedules, it is apparent that in many if not most hospitals, rotations schedules are known to the service chiefs but not to anyone in central administration. Moreover, central administration staff encountered considerable resistance as they tried to collect FTE data, and I imagine some of that resistance translated into poor data.[2]

Resident-to-Bed Ratios. From the Rand Corporation, we received data on resident-to-bed ratios for New York State

[2] C. Dichter, Bureau of Health Economics, Office of Health Systems Management, Department of Health: Letter to Medical Practice Study (Feb 10, 1987)

hospitals in 1984 and 1986. Many of the data were missing for one or both years. Although we could not use these ratios as the sole indicator of teaching status, we were able to use the data to reflect the involvement of residents in patient care.

II. Results of Data Analyses

Table 1 shows that measures of teaching status are inconsistent. Of the 79 teachings hospitals as defined by the SEDI index, 6 (8%) lacked approved residency programs according to the AHA data although they all had positive residency-to-bed ratios. One hospital facility with a large number of approved residency positions had no data for the 1984 or 1986 resident-to-bed ratios. Thirty-one non-SEDI hospitals reported approved residency programs. Five of those 31 had fewer than .5 FTEs per 1,000 discharges. For many others the residency programs were limited, both in proportion to number of beds and number of discharges.

One hospital reported the highest resident-to-bed ratio of all 270 hospitals but had no approved positions. This inconsistency is likely an error in data.

The 13 facilities designated the university teaching hospitals of the state's medical schools had wide-ranging resident-to-bed ratios (.212 to .585) but a common and significant integration with an academic institution.

Five specialty hospitals not included in the 1984 SEDI teaching category warranted special consideration. All had (1) approved residency programs as listed by AHA and (2) resident-to-bed, FTE-to-bed, and FTE-per-1,000 discharge ratios that were more like those of the SEDI teaching hospitals than those of the SEDI non-teaching hospitals. Since 1984, Children's Hospital has been reclassified as a SEDI teaching facility.

III. Discussion

Using these data we chose the following definitions for sampling and analyses.

First, for the 13 SEDI teaching hospitals designated by medical schools as primary teaching hospitals, we established the classification university teaching hospital.

Next, we chose to adopt, with the modifications noted below, the Department of Health's SEDI index as the principal determinant for classification of all other teaching facilities (affiliate teaching hospitals). Many of the SEDI non-teaching hospitals with AHA listings for residency programs appeared to have limited commitment to training; therefore, we could not justify treating them as teaching facilities.

We decided to include in this category of affiliate teaching the five specialty hospitals named above which were not in the SEDI index as teaching facilities. As explained in the main text of Chapter 4, for sampling purposes these specialty institutions were all clustered with general care teaching institutions and therefore appeared in the substrata for teaching hospitals.

Finally, all remaining hospitals were designated non-teaching.

For sampling, we classified hospitals into teaching and non-teaching only; we did not stratify university teaching facilities separately. Analyses in Chapter 6 use the 3-level indicator for the facilities selected for the Study.

TECHNICAL APPENDIX 4.II.1

TABLE 1 Teaching Hospitals--Sorted by SEDI and Resident/Bed Ratios

No.	HOSPITAL NAME	SPECIALTY	Medical School Primary Hosp	SEDI	AHA3	AHA5	Res/Bed 1984 (or 86)	FTEs per bed	FTEs per 1000 dchg
1	LIJ-MANHASSET		.	1	1	1	.594	.362	9
2	LIJ-QUEENS		.	1	1	1	.594	.276	9
3	LIJ-SCHNEIDERS	CHILDREN		1	1	1	.594	.	10
4	U HOSP-STONY BROOK		1	1	1	1	.585	.752	22
5	ALBANY MED CTR		1	1	1	1	.516	.375	11
6	NORTH SHORE UNIV H		.	1	1	1	.466	.383	7
7	STATE U UPSTATE MC		1	1	1	1	.433	.532	19
8	NASSAU CTY MC		.	1	1	1	.412	.423	11
9	BETH ISRAEL HOSPITAL		.	1	1	1	.410	.283	8
10	MONTEFIORE MC		1	1	1	1	.389	.232	9
11	MONTEFIORE-EINSTEIN		1	1	1	1	.389	.387	9
12	WESTCHESTER CO CTR		1	1	1	1	.379	.160	6
13	LINCOLN MEDICAL CTR		.	1	1	1	.361	.528	10
14	BROOKDALE MED CTR		.	1	1	1	.356	.334	8
15	MOUNT SINAI HOSPITAL		1	1	1	1	.355	.405	13
16	BRONX MUNICIPAL HOSP		.	1	.	.	.352	.600	21
17	BRONX-LEBANON FULTON		.	1	1	1	.340	.281	8
18	BRONX-LEBANON CONC		.	1	1	1	.340	.406	8
19	ST LUKES ROOSEVELT		.	1	1	1	.333	.338	9
20	ST-LUKES HOSPITAL		.	1	1	1	.333	.295	9
21	N CENTRAL BRONX		.	1	1	1	.327	.334	8
22	MEMORIAL H CANCER	CANCER		1	1	1	.322	.459	15
23	ST VINCENTS MED CTR		.	1	1	1	.322	.170	6
24	PRESBYTERIAN H (NYC)		1	1	1	1	.317	.444	13
25	NEW YORK H-CORNELL		1	1	1	1	.316	.413	11
26	MARY I BASSETT HOSP	RURAL TEACHING		1	1	1	.311	.276	8
27	CONEY ISLAND HOSP		.	1	1	1	.303	.338	9
28	HARLEM HOSPITAL CTR		.	1	1	1	.298	.315	11
29	INTERFAITH MED CTR		.	1	1	1	.292	.077	5
30	ST JOHN EPISCOP DIV		.	1	1	1	.292	.221	5
31	METHODIST HOSP BRKLN		.	1	1	1	.284	.261	7
32	HIGHLAND HOSPITAL		.	1	1	1	.280	.225	4
33	BOOTH MEMORIAL		.	1	1	1	.278	.257	6
34	METROPOLITAN HOSP		.	1	1	1	.278	.444	15
35	LONG ISLAND COLLEGE		.	1	1	1	.266	.256	7
36	WOODHULL MED CTR		.	1			.257	.351	8
37	CITY HOSP ELMHURST		.	1	1	1	.253	.312	11
38	MAIMONIDES MED CTR		.	1	1	1	.249	.306	8
39	DOWNSTATE MED CTR		1	1	1	1	.247	.580	15
40	KINGS COUNTY HOSP		.	1	1	1	.241	.326	10
41	LENOX HILL		.	1	1	1	.235	.237	6
42	JAMAICA HOSP		.	1	1	.	.223	.213	6
43	QUEENS HOSPITAL		.	1	1	.	.222	.213	7
44	NYU MEDICAL CENTER		1	1	1	1	.222	.317	11
45	ERIE COUNTY MED CTR		1	1	1	1	.219	.295	11
46	STRONG MEMORIAL H		1	1	1	1	.212	.366	10
47	CABRINI MED CTR		.	1	1	1	.205	.190	6
48	KINGSBROOK JEWISH		.	1	1	1	.199	.177	6
49	ROCHESTER GENERAL H		.	1	1	1	.198	.120	3
50	FLUSHING HOSP MCTR		.	1	1	1	.193	.182	4
51	BROOKLYN CALEDONIAN		.	1	1	1	.190	.257	6
52	CALEDONIAN DIV		.	1	1	1	.190	.201	6

TECHNICAL APPENDIX 4.II.1

No.	HOSPITAL NAME	SPECIALTY	Med School Primary Hosp	SEDI	AHA3	AHA5	Res/Bed 1984 (or 86)	FTEs per bed	FTEs per 1000 dchg
53	WINTROP U/NASSAU		.	1	1	1	.189	.218	5
54	ST JOSEPH HEALTH C		.	1	1	1	.189	.127	3
55	STATEN ISLAND HOSP		.	1	1	1	.187	.178	4
56	LUTHERAN MED CTR		.	1	1	1	.186	.176	5
57	BUFFALO GENERAL H		.	1	1	1	.181	.174	6
58	BGH DEACONESS DIV		.	1	1	1	.181	.132	6
59	ST MARYS H ROCHESTER		.	1	1	1	.177	.121	4
60	ST VINCENTS H RICHMD		.	1	1	1	.172	.216	5
61	BELLEVUE HOSPITAL		.	1	1	1	.165	.104	5
62	GENESSEE H ROCHESTER		.	1	1	1	.162	.102	2
63	WYCOFF HEIGHTS		.	1	1	1	.162	.110	4
64	CMC-ST.MARY'S BKLYN		.	1			.135	.000	0
65	BAYLEY SETON HOSP		.	1	1	.	.135	.066	2
66	CMC-ST. JOHN		.	1	1	.	.133	.277	7
67	CMC-HOLY FAMILY		.	1	1	.	.133	.134	6
68	CMC-MARY IMMACULATE		.	1	1	.	.133	.227	7
69	UHS-BINGHAMTON GENL		.	1	.	.	.121	.041	1
70	UHS-IDEAL DIV		.	1	.	.	.121	.009	1
71	UHS-WILSON DIV		.	1	.	.	.121	.056	1
72	HIP LAGUARDIA		.	1	1	.	.121	.146	3
73	MILLARD FILMORE HOSP		.	1	1	1	.113	.102	3
74	FILLMORE SUBURBAN		.	1	1	1	.113	.212	3
75	S. NASSAU COMMUNIT.		.	1	1	1	.060	.051	1
76	ST PETERS HOSPITAL		.	1	1	1	.012	.056	1
77	COMM GEN OF SYRACUSE		.	1	1	1	.007	.012	0
78	CROUSE-IRVING MEM		.	1	1	1	.000	.141	3
79	ROSWELL PARK	CANCER	.	1	1	1	.	1.24	41
80	WESTCHESTER SQ MC		.	.	.	.	.698	.000	0
81	OUR LADY OF MERCY		.	.	1	1	.231	.263	7
82	HOSP JOINT DISEASES	ORTHOPEDICS	.	.	1	1	.222	.266	8
83	MANHATTAN EE & T	ENT	.	.	1	1	.213	.244	3
84	NY EYE & EAR INF	ENT	.	.	1	1	.209	.239	3
85	NY INF BEEKMAN		.	.	1	1	.203	.199	6
86	SHEEHAN EMERGENCY H		.	.	1	.	.184	.000	0
87	JOINT DISEASES N GEN		.	.	1	1	.145	.098	3
88	MERCY HOSP BUFFALO		.	.	1	1	.138	.089	2
89	NEW ROCHELLE H MC		.	.	1	1	.138	.098	3
90	BAPTIST MED CTR		.	.	.	.	.130	.130	4
91	HOSP SPECIAL SURGERY	ORTHOPEDICS	.	.	1	1	.125	.219	9
92	ST JOSEPHS H YONKERS		.	.	1	1	.119	.090	3
93	ST CLARES HOSPITAL		.	.	1	1	.078	.073	2
94	SISTERS OF CHARITY		.	.	1	1	.073	.067	2
95	ST CLARES HEALTH CTR		.	.	.	.	.071	.006	0
96	MT VERNON HOSPITAL		.	.	1	1	.057	.099	3
97	COMMUNITY H GLENCOVE		.	.	1	1	.057	.024	1
98	BROOKHAVEN MEM HOSP		.	.	.	.	.052	.051	1
99	MERCY HOSP ROCKVILLE		.	.	1	1	.050	.025	1
100	KINGSTON HOSPITAL		.	.	1	1	.047	.039	1
101	SOUTHSIDE HOSPITAL		.	.	1	1	.047	.049	1
102	PENINSULA HOSP CENTER		.	.	.	.	.046	.053	2
103	LONG BEACH MEM H		.	.	.	.	.041	.027	1
104	NIAGRA FALLS MEM MC		.	.	.	.	.028	.028	1
105	WHITE PLAINS H MC		.	.	.	.	.026	.020	0
106	HUNTINGTON HOSPITAL		.	.	1	1	.015	.012	0
107	ST CHARLES H- PORT J		.	.	1	1	.015	.008	0
108	BENEDICTINE HOSPITAL		.	.	1	1	.014	.027	1

No.	HOSPITAL NAME	SPECIALTY	Med School Primary Hosp	SEDI	AHA3	AHA5	Res/Bed 1984 (or 86)	FTEs per bed	FTEs per 1000 dchg
109	CHAMPLAIN VALLEY MC		.	.	.	.	.011	.011	0
110	ELLIS HOSPITAL		.	.	.	1	1	.010	.032	1
111	NYACK HOSPITAL		.	.	.	.	.	.009	.003	0
112	MASSAPEQUA GENERAL		.	.	.	.	.	.008	.061	1
113	ST ELIZABETH HOSP		.	.	.	1	1	.008	.056	2
114	CHILDS HOSPITAL		.	.	.	1	1	.000	.081	1
115	HEMPSTEAD GENERAL		.	.	.	.	.	.000	.018	0
116	ST. FRANCIS	CARDIOVASCULAR	.	.	.	.	.	.000	.053	2
117	GOOD SAMARITAN SUFF		.	.	.	.	.	.000	.003	0
118	G SAMARITAN W. ISLIP		.	.	.	1	.	.000	.002	0
119	PARKWAY HOSPITAL		.	.	.	.	.	.000	.000	0
120	RICHMOND MEM HOSP		.	.	.	.	.	.000	.005	0
121	CMC-ST JOE(HLCR-OST)		.	.	.	1	1	.	.000	0
122	CHILDRENS H BUFFALO	CHILDREN		.	.	1	1	.	.355	4
123	SUNNYVIEW H & REHAB	OTHER		.	.	1	1	.	.020	2
124	CHURCH CHAR-SMITHTOWN		.	.	.	1	1	.	.000	0
125	ST BARNABAS HOSPITAL		.	.	.	1	1	.	.106	4
126	CHURCH CHAR-S SHORE		.	.	.	1	1	.	.000	0

TABLE 2 HOSPITALS DESIGNATED AS TEACHING BY SEDI INDEX

N Obs	Label	N	Minimum	Maximum	Nmiss	Mean
79	Res/Bed 84	78	0.000	0.594	1	0.259
	FTEs per bed	78	0.000	1.242	1	0.268
	FTEs/ 1000 disch	79	0.000	40.964	0	7.725

TABLE 3 HOSPITALS DESIGNATED AS NON-TEACHING BY SEDI INDEX

N Obs	Label	N	Minimum	Maximum	Nmiss	Mean
47	Res/Bed 84	41	0.000	0.698	6	0.086
	FTEs per bed	47	0.000	0.355	0	0.071
	FTEs/ 1000 disch	47	0.000	8.730	0	1.801

TABLE 4 ALL POSSIBLE TEACHING HOSPITALS

N Obs	Label	N	Minimum	Maximum	Nmiss	Mean
126	Res/Bed 84	119	0.000	0.698	7	0.199
	FTEs per bed	125	0.000	1.242	1	0.194
	FTEs/ 1000 disch	126	0.000	40.964	0	5.515

ADJUSTMENT FOR PROBABILITY OF SELECTING SMALL HOSPITALS

This appendix outlines the weighting adjustments used to compensate for the method we used to group undersized primary sampling units (PSUs) with additional hospitals. The purpose of grouping was to eliminate the gross inefficiency associated with high sampling fractions and to avoid the resulting confidentiality problems. Grouping reduced the probability of selecting small hospitals, but the following adjustments to 18 of the 51 hospitals in the sample compensated for the reduced probability.

These calculations are based on the methods used to select hospitals systematically from the sampling frame (see main text). In most cases, the hospitals in our sample could be selected only in the manner outlined in the text (i.e., when the hospital's cumulative sampling units (CSU) coincided with the selection number). Most hospitals were sufficiently large (8,000 discharges per year) that they would not be clustered with another facility.

On the other hand, some hospitals could also be selected because they were joined with undersized hospitals that were selected as above. In these situations, a single facility had more than one chance of being selected, depending on its proximity on the sampling frame with undersized units to which it might be joined. For example, in the sampling frame in Table 1, if the random selection number was 60, hospital C would be selected directly with a probability of 20/86. On the other hand, if the random selection number was 65, hospital D would be selected directly; but because it is undersized, hospital C would be clustered with it and selected indirectly. In this case, the probability of selecting hospital C indirectly is 6/86.

TABLE 1

Hospital	Sampling Units	CSU
A	30	30
B	12	42
C	20	62
D	6	68
E	10	78
F	8	86

In general, the probability of selecting a specific record within any hospital, H, is a sum of probabilities because the hospital could be selected in more than one way. The probability of selection of a record within a hospital or cluster depends on the number of sampling units within the cluster. Let R_{ij}= the ith record in hospital j, then the probability of selection of that record is:

$$p(R_{ij}) = \sum_{k=1}^{K} P(R_{ijk} | H_{jk}) * P(H_{jk})$$

where k=1,...,K are the different ways in which hospital j can fall into the sample using our sampling rules.

Applying this general formula to the sampled hospitals, the weight assigned to an observation is the reciprocal of its probability of selection. Without considering undersized sampling units or over/under sampling within a hospital, the probability of selecting any record would be:

p = (SU/86)*(1/SU) = 1/86, where:

SU = the number of sampling units in a hospital or cluster,

SU/86 = the probability of selecting a given cluster, and

1/SU = the probability of selecting a specific record within the cluster.

Adjusting for the over/under sampling of records within a cluster:

$$p = (SU/86)*(1/SU)*(1/STR_m)$$

in which STR_m = the over/under sampling factor for the m^{th} stratum: 2 for the elderly 1/3 for high-risk non-elderly, and 1 for all others.

Next, correction of sampling frequencies was needed to obtain the desired overall number of records and at the same time to maintain the relative numbers of oversampled, undersampled, and base-rate sampled records. Therefore, prior to adjusting for undersized hospitals, the probability of selection of a record was:

$p = (SU/86)*(1/SU)*(1/STR_m)*(1/c)$ where

c = the adjustment constant (1.1, for ensuring about 31,000 records in the sample).

Finally, an adjustment for multiple methods of selecting a hospital would accommodate the undersized facilities. Using this adjustment w_j for hospital j, then the probability of selection is:

$$p = (SU/86)*(1/SU)*(1/STR_m)*(1/c)*(1/w_j)$$

and the weight assigned to the record is:

$wt=1/p$

Table 2 presents the adjustments for 18 hospitals to account for undersized units. The remaining hospitals need no adjustment (i.e., $w_j = 1$). This adjustment, wj, changed the weights of approximately 30% of the 31,429 records in the sample. Of the weights that were adjusted, 20.0% were above or below the

unadjusted high and low weights of 31.53 and 189.2.

These adjustments reduce the possibility of bias due to under-representation of small hospitals.

TABLE 2

Adjustment of case weights for hospitals in sample owing to undersized units in the sampling frame

Hospital Number	Possible Methods of Selection	Prob Select* Hosp	Prob Select Record	Selection probability (86=base)	Weight adjustment
1	1	(4/86)	*1/10		
	2	(7/86)	*1/11	82.982	0.9649123
2	1	(7/86)	*1/11	135.143	1.5714286
3	1	(3/86)	*1/9		
	2	(3/86)	*1/13	152.455	1.7727273
4	1	(6/86)	*1/8		
	2	(3/86)	*1/9	79.385	0.9230769
5	1	(1/86)	*1/8		
	2	(1/86)	*1/9		
	3	(1/86)	*1/16	288.000	3.3488372
6	1	(1/86)	*1/16		
	2	(1/86)	*1/8	458.667	5.3333333
7	1	(14/86)	*1/14		
	2	(1/86)	*1/16		
	3	(5/86)	*1/19	64.873	0.7543424
8	1	(8/86)	*1/8		
	2	(5/86)	*1/13	62.111	0.7222222
9	1	(5/86)	*1/13		
	2	(2/86)	*1/14	163.042	1.8958333
10	1	(1/86)	*1/10		
	2	(3/86)	*1/10	215.000	2.5

11	1	(3/86)*1/10		
	2	(1/86)*1/10		
	3	(1/86)*1/9		
			168.261	1.9565217
12	1	(1/86)*1/9		
	2	(1/86)*1/10		
	3	(3/86)*1/10		
	4	(5/86)*1/14		
			99.049	1.1517367
13	1	(5/86)*1/14		
	2	(1/86)*1/10		
	3	(3/86)*1/10		
	4	(1/86)*1/9		
			99.049	1.1517367
14	1	(12/86)*1/12		
	2	(5/86)*1/17		
			66.455	0.7727273
15	1	(13/86)*1/13		
	2	(1/86)*1/14		
			80.267	0.9333333
16	1	(12/86)*1/12		
	2	(6/86)*1/18		
			64.500	0.75
17	1	(12/86)*1/12		
	2	(3/86)*1/15		
			71.667	0.8333333
18	1	(10/86)*1/10		
	2	(1/86)*1/11		
	3	(1/86)*1/14		
			73.989	0.8603352

[a] First, a hospital can be selected directly; second, it can be selected indirectly after being joined with another hospital. Indirect selection can occur in more than one manner.

[b] Selection probability of a record for each of the K ways in which it can be sampled (assuming no over/under sampling).

[c] Reciprocal of the adjusted selection probability for a record in the given hospital (e.g., 86 means 1/86).

[d] Weight adjustment, w_j, assigned to any record in that hospital. For

remaining 33 hospitals, the factor $w_j = 1.0$.

Chapter 5

THE NATURE AND INCIDENCE OF MEDICAL INJURY: METHODS FOR MEDICAL RECORD REVIEW

I. Introduction

The medical record review was key to estimating the incidence of adverse events associated with medical management. By means of this review, the Study identified those patients who suffered disability because of health care and the subset of patients whose injuries were caused by negligence. Thus, the record review focused on two critical issues: causation and negligence. We asked, "Was the patient's condition attributable to medical management rather than to the disease under treatment (causation)? Was negligence involved?"

In addition to establishing causation and negligence, the record review sought to describe the environment in which the injury was sustained (e.g., the facility where the injury occurred, specialty involved, type of medical care rendered). Reviewers then classified the injury using a clinical methodology and rated it according to the magnitude of disability experienced.

Design of the medical record review in part replicated, and in part greatly expanded upon the California Medical Association (CMA) study conducted in California in the mid 1970s.[1] The review was conducted by teams of trained medical records administrators (MRAs) and nurses (for the screening phase) and board-certified physicians (for the physician review phase). Each phase presented its own unique and specific logistic and methodologic problems. The review presented logistic problems because the infrequency of adverse events necessitated design of efficient methods for sifting through thousands of medical records to find the few that indicated patient disability caused by medical management. The methodologic problems evolved

---

[1] California Medical Association, Report on Medical Insurance Feasibility Study (Sutter Publications 1977)

because MRAs had to make valid judgments regarding the presence of screening criteria and physicians had to make valid and reliable judgments about whether a patient's injury resulted at least in part from medical management and, if so, whether management failed to meet the standard of medical care.

In this chapter, we describe the methods we employed to resolve both the logistic and methodologic issues of the medical record review: (a) the manner in which we went about securing the cooperation of 51 hospitals; (b) the detailed mechanics of the record review in New York State; (c) a number of issues regarding concepts of causation and negligence; (d) the reliability and validity of methods we used for making these judgments. Finally, we present an overview of our methods for analyzing the adverse event data.

II. Review of Records in New York State

A. Hospital Recruitment

Success of the medical record review demanded active cooperation of sample hospitals. In July 1987, Commissioner of Health Dr. David Axelrod wrote to the 51 sample hospitals to announce the Study and solicit participation. After that mailing, the Department of Health (DOH) sought hospital participation through the New York State Hospital Association and the Hospital Trustees Association.[2] In November 1987, the Department forwarded to the Study office a complete list of hospitals that had thus far responded to the Commissioner's letter. Following receipt of this list of approximately half of the sample hospitals, the Study sent a draft of its proposed contact letter to the DOH for review and approval. The Commissioner approved the Study's contact letter and authorized

[2] In early 1987 we completed a pilot study in three New York Hospitals (hereinafter, the Pilot Study). In 1987-8, we completed pilot studies of methodological issues in Massachusetts (hereinafter, the Intermediate Pilot Studies).

the Medical Practice Study to contact hospitals on December 28, 1987.

In January 1988, the Medical Practice Study began contacting the selected hospitals to solicit their approval of, and participation in, the Study. The initial letter described the project in detail, specified the time frame of events, and asked the hospitals to release their SPARCS[3] data at the DOH to facilitate execution of the sampling plan (see Appendix 5A).

By March 7, 1988, 48 of the 51 sampled hospitals had agreed to participate in the Study. On March 16, 1988, the Study began receiving SPARCS data tapes from the DOH. From these tapes we drew the sample of hospitalizations for review. Actual record review began in May, 1988. By late summer all 51 hospitals had agreed to participate and sent in SPARCS release letters.

A few hospitals were reluctant to participate. In some cases the reluctance stemmed from a misunderstanding of the role of the Study in relation to existing DOH regulatory activities. In other cases, the hospitals voiced concerns about the confidentiality of results and the risks of litigation. The Study office endeavored to allay these fears. Several institutions requested Study participants to visit their facilities and meet with hospital staff to address their concerns about the project. We responded to all such requests.

Because of the sensitive nature of medical record review, the participating hospitals and their physicians demonstrated concern over the release of information. The Medical Practice Study was conducted under the authority of Public Health Law §206(1)(j). Therefore, patient-, physician-, and hospital-specific results are confidential. They are not available to the Commissioner of Health or others. As a condition of working on this project, all persons were required to sign a statement of understanding of confidentiality and not to release any patient-,

[3] Statewide Planning and Research Cooperative System. See Chapter 4.

physician-, or hospital-specific information. All of the Study's internal files carry a unique identifier to protect the identity of patients and institutions. Thus, we have been able to assure all hospitals that their participation will not jeopardize confidentiality of their patient records or subject them to penalty for full disclosure.

B. Record Review Process

The Study hospitals were divided into seven geographic zones, three in New York City, one in Long Island, and the remaining three in upstate New York. In charge of each zone was a Study physician-supervisor, who was a member of the Study staff in Boston, assisted by an MRA-supervisor. MRA-supervisors were administrators recruited by the Hospital Association of New York State (HANYS) with experience in conducting research projects in record rooms. The MRA-supervisor provided ongoing liaison with the record room staff at each hospital. We asked the hospitals to provide all medical records in the sample in their entirety.

Our two-phase design required coordination of the screening process undertaken by the MRA staff with the physician review of screened records. After the MRAs screened records, the physician-supervisor led a team of twelve to fifteen physician-reviewers in his particular zone for the second level record review. All physician-supervisors had extensive, practical training in the review of records for adverse events. Five of them had developed the adverse event analysis form (AEAF) and its associated documentation. They maintained constant contact with the MRA-supervisor at each hospital, and both physician-supervisors and MRA-supervisors ensured that physician-reviewers were receiving the screened records in an appropriate fashion.

Several physician specialty societies and medical schools in New York helped us recruit physician-reviewers, 96% of whom were board-certified (70% internists, 22% general surgeons). Each physician-reviewer received comprehensive instructional materials

and a two-hour training session conducted by the physician-supervisor. All physicians and MRAs executed a statement of confidentiality of the information in the medical record review. We maintained records establishing the identity of each screener and reviewer for all records to permit us to monitor their performance.

The day-to-day aspects of the screening process were carefully overseen by the MRA-supervisors. We trained the MRAs in the use of our screening criteria in a two-hour session conducted jointly by HANYS personnel and the Study staff. The MRA-supervisors ensured that the screening process was done properly and answered questions of the MRAs during the screening phase. In addition, as described in a following section of this chapter, the MRA supervisor repeated reviews of a sample of 1% of all the records to check the validity of the MRA screens.

The physicians assessed the medical records that contained positive screening criteria for evidence of adverse events and negligence. After reviewing the MRA screening form and the medical record itself, they first determined whether a screening criterion had been met and if there was a possibility of an adverse event. If they found any evidence of an adverse event, they graded their confidence in its presence on a scale of 1 to 6. We refer to this as the causation score because it reflects the reviewer's confidence in his or her judgment that medical management(MM), rather than the disease process, caused the adverse event. If the confidence score for causation was greater than one, the reviewer assessed the disability caused by the adverse event as well as certain clinical features of the adverse event. We used these data to assess the severity of the adverse event and to provide a comparison with self-reports by patients.

Physician-reviewers originally used a nine grade disability scale that was subsequently collapsed to the seven grade scale (Table 5.1).

Table 5.1
Description of Disability Class

Disability Class	Description
One	Minimal Impairment with Complete Recovery in One Month
Two	Moderate Impairment Requiring One to Six Months for Recovery
Three	Moderate Impairment Requiring More Than Six Months for Recovery
Four	Permanent Impairment Involving 1-50% Decrease in Social Functioning
Five	Permanent Impairment Involving >50% Decrease in Social Functioning
Six	Death
Seven	Cannot Reasonably Judge Disability

Physicians had to base their decisions about disability on evidence available in the medical record, including any hospitalizations subsequent to the index one. Without complete follow-up information on the patient, error-free judgments of disability are impossible. Moreover, physician judgments that an adverse event led to death must be understood in light of possibly extenuating factors. For example, many elderly and/or critically ill patients who died as a result of an adverse event likely would have died of their underlying disease within a short time, but we did not ask physicians to estimate the number of days of life lost. Although judgments regarding critically ill patients would be difficult in any event, some patients may have requested and received limited care without that fact being documented in the medical record, making the situation even more complicated. While we trained physician-reviewers to be alert to such issues, they may still have led to errors in our estimates.

After estimating disability, physician-reviewers judged whether the evidence indicated negligence and their confidence in that judgment. They also classified the nature of the negligence and graded its severity in terms of its deviation from the

standard of care for 1984. During the review process, the physician-reviewers could consult experts in neonatology, orthopedics, neurosurgery and obstetrics who were recruited by the Study from faculties of New York medical schools. In addition, the Study office in Boston provided access to consultants in other fields. Although pilot studies had shown physician judgments can be valid without expert input, we nonetheless thought that access to experts would prove helpful in certain cases.

Since we were interested in defining the universe of adverse events, physicians-reviewers recorded not only adverse events that arose in, and were discovered during, the index hospitalization but also those adverse events caused by medical management in the index hospitalization but not discovered until later, and those caused earlier that led to, or were first discovered during the index hospitalization. They also recorded the date of the medical management that caused the adverse event if it did not occur in the index hospitalization.

Each record screened as positive was subjected to two independent physician reviews. After both physician reviews were completed, the MRA-supervisor reviewed the AEAFs to assure that the two physician-reviewers had judged the same adverse event. If any discrepancy in the description, timing, or discovery of the adverse event existed between the two AEAFs, the MRA-supervisor covered over the scores for negligence and adverse event causation and referred the forms to a physician-supervisor. The physician-supervisor reviewed the discrepant AEAFs and the medical record and then resolved the discrepancy by changing an obvious oversight or minor mistake in dates on one of the two AEAFs or by completing an independent review, thereby replacing one AEAF or, rarely, both of the previous AEAFs. Using this information, we made a determination of adverse event and negligence status for each case as described in Section V.B.

After completion of the record review, we classified each adverse event using information from the AEAF. Each adverse event was classified by one of two Study physician-supervisors, one a surgeon, the other an internist, after reading the AEAFs completed by both physician-reviewers for each case. Problem cases were discussed and classified by consensus. In the vast majority of instances, including the major categories of drug reactions, wound infections, procedure-related injuries, and diagnostic and therapeutic mishaps, the classification was obvious. This step completed the data collection for our study of medical records.

We realized throughout data collection that our record review was based on two critical concepts: causation of adverse events and negligent care. We believed it was important to define these concepts in some detail. Moreover, we felt that the integrity of the record review would be related directly to our ability to make valid and reliable judgments of causation and negligence. In the following sections, we describe the concepts of causation and negligence in detail and then present data regarding the reliability and validity of our judgments.

III. Conceptual Issues Surrounding Record Review

The medical record review process involved a series of difficult conceptual issues, the most important of which concerned judgments of causation and negligence. In this section we describe the critical conceptual issues that underlie judgments of causation and negligence as well as issues arising in the estimate of incidence. In the following section we discuss the validity and reliability of our process for identifying adverse events and negligence.

A. Judgments of Causation

Our initial experiences in the Pilot Study focused our attention on the importance of judgments about causation of injury and negligent behavior by physicians. These issues have long been debated and discussed in the jurisprudence of tort law.[4] Tort law compensates only for those injuries caused by medical management provided in a substandard or negligent fashion whereas a no-fault system would compensate all injuries caused by medical management. Since the law generally distinguishes findings of causation from those of negligence, we do so as well in this discussion.

Both tort and no-fault systems must contend with the problem of identifying the particular event that can be said to have caused an injury. Many injuries have several causes and the problem of singling out one particular cause is troubling.[5] Problems of causation are particularly important in malpractice law. The malpractice litigant must demonstrate an injury that arose from medical management. The plaintiff is generally suffering from one or more diseases, however, and frequently the injury is related to an underlying disease. Therefore, courts face the highly technical issues underlying the relationship between the putatively iatrogenic injury and the medical management.

Usually, disentangling the adverse event from the disease process is difficult. Suppose, for example, that someone suffered a badly injured leg in an auto accident, was hospitalized for treatment, and eventually left the hospital with a permanent limp. The costs of this disability are properly chargeable to the health care system if, and only if, the

---

[4] Brennan, Causal Chains and Statistical Links: The Role of Scientific Uncertainty in Hazardous Substance Litigation, 73 Cornell L. Rev. 469-533 (1988)

[5] Calabresi, Concerning Cause and the Law of Torts: An Essay for Harry Kalven, Jr., 43 U. Chi. L. Rev. 69 (1975).

disability was actually caused by the treatment rather than by the injury suffered in the accident. Disentangling the consequences of medical intervention itself from the consequences of the original condition requiring treatment is difficult in a case such as this.

Actually, even finding that a particular disability was the consequence of medical treatment is not sufficient: the injury must be the unintended consequence of such treatment. As an illustration of the difference, suppose that the leg had been seriously injured in the accident and then had also become infected. The patient had not sought therapy for some time, and the leg had become gangrenous. When the patient came to the hospital, the judgment was made that amputation of the leg was required to save the patient's life. Here it is true to say, as a factual matter, that the treatment was a cause of loss of the leg. As a matter of compensation policy, though, one would not conclude that amputation was the relevant cause because the initial unhealthy condition made this mode of treatment desirable and thus was the real cause of the loss of the leg. What we sought in the medical record survey, then, are the unintended, adverse consequences of management, not the intended consequences of an appropriate intervention designed to benefit the initial condition.

A further complication arises in applying this causal standard. The example above described positive consequences of active intervention or acts of commission. A different problem is posed by negative consequences of failure to intervene or acts of omission. Suppose, as illustration, that a patient undergoing treatment for cancer dies in the hospital as a result of the cancer. The death was caused by the original condition, not the treatment. The events would be viewed differently if the patient had died from severe toxicity caused by an inadvertent overdose of chemotherapy. The latter patient would be said to have suffered injury secondary to medical management. Further, if the

cancer was curable (e.g., early Hodgkins lymphoma) but was not diagnosed and treated properly in the hospital, the death resulting from the lymphoma would actually be attributable to an avoidable medical failure, not the original cancer. Thus, the loss of months or years of life, secondary to the failure to diagnose a treatable disease, is caused by medical management. In other words, failures to make diagnoses or to employ proper therapy, or errors of omission, result in injuries attributable to medical management.

Once one grants that conclusion, another issue arises. What is the criterion for determining what should have happened? Judgments of errors of omission require setting standards for what should have been proper diagnosis and treatment. Among the easy cases are those in which the proper diagnosis was not made but for which treatment was impossible. If a disease cannot be cured, one cannot find it to be the responsibility of any compensation program, including that governed by tort law, to pay for the losses suffered from the disease. Suppose, though, that the disease could have been cured if it had been treated appropriately. In this case it is correct to say that medical management caused the loss of years or months of life. But determination of the extent of the injury depends on one's judgment regarding when the diagnosis should have been made, a judgment that necessarily involves an assessment of the standard of care. Thus, considerations of standard of care, or negligence, are not completely separable from judgments of causation, especially in acts of omission.

Removing medical accident compensation from a tort law framework does not do away with problems of causation. No-fault systems also compensate according to causation but do not have to determine negligence in most cases, making overall decisions about compensation somewhat less complicated. Nonetheless, those jurisdictions that employ a no-fault approach to medical injury are still going to be troubled by the generic problems with the

causal relationship between accident and disease as outlined above. In a no-fault system, negligence does not need to be demonstrated, but an individual claiming compensation must still prove that the injury for which he seeks compensation was caused by an accident in medical management rather than by his disease. Thus, even no-fault jurisdictions must struggle with the causal problems noted above: disentangling medical management from disease processes (including distinguishing between the unavoidable repercussions of a disease and those that could have been avoided if the disease had been treated appropriately or diagnosed sooner), and the distinction between the intended results of medical intervention and adverse consequences of medical management.

Some cases from New Zealand and Sweden, which have no-fault administrative schemes for medical accident compensation, illustrate the three kinds of generic problems that arise in attributing causation. In the case of the Estate of Hellaby,[6] an elderly woman died 18 days after a fall in a hospital that resulted in a fracture. In the initial report the pathologist suggested that the death was related to old age. In a later report, however, he certified that a significant contribution to death had been pneumonia that could have resulted from immobilization caused by the fracture. Based on this information, New Zealand authorities decided that a causal link existed between the fracture and the death and so provided compensation. They thus disentangled the background process of aging from the repercussions of the accident.

A similar case is Viggars v. Accident Compensation Corporation,[7] in which a man suffered several small strokes. To evaluate the cause of these strokes, he underwent carotid

---

[6] 1 NZAR 344 (1977)

[7] 6 NZAR 235 (1986)

angiography, a procedure in which a catheter is introduced into certain neck blood vessels. During the procedure, the patient suffered another stroke. Medical testimony was that the catheter itself could have dislodged some clot from the vessel wall and that this clot could then have traveled to the brain and caused the stroke. Other physicians testified that the stroke could have been caused by the same disease process that caused the first stroke. Given the temporal contiguity and the greater likelihood of the former explanation, the court found for the petitioner and granted compensation.

In another case, Cutler v. Vauxhall Motors Ltd,[8] the patient suffered a minor ankle injury that led to an ulcer. The physician discovered, however, that the plaintiff had severe varicose veins in both legs, a condition that pre-dated the injury and that can lead to such ulcers. An operation ultimately had to be performed to repair the ulcer, but the court would not allow full compensation because the pre-existing condition allowed a minor accident to become a major one. In each of these cases, the court had to disentangle the disease process from medical management.

Other cases illustrate the second kind of problem with causation, the fine line between intended and adverse outcomes of medical therapy. For example, chemotherapy for treatment of cancers may lead to low white blood cell levels. In part because diminished white blood cells decrease the ability to fight bacteria, many patients develop infections. The intended outcome, destruction of tumor cells, is thought to outweigh the adverse outcome, infections caused by loss of white blood cells. Thus, injuries caused by infections in patients treated with chemotherapy are not compensated in New Zealand[9] or Sweden.[10]

---

[8] 1 QB 418 (1971)

[9] Deutsch, Medical Malpractice and Medical Misadventure in New Zealand: Public Insurance in Lieu of Private Liability as Administered by the Courts and the Accident Compensation Commission. 1 Med L. 345-54 (1982)

No-fault regimes have also addressed problems of distinguishing unavoidable outcomes of diseases from failed therapies or failures to make diagnoses. Consider, for example, Re Carroll: Decision No. 59/84.[11] The claimant, presenting with severe abdominal pain, was thought to have appendicitis and was taken to surgery. The appendix was not inflamed and no other diagnosis was made. The pains persisted, and the patient died three days later. On postmortem examination, a leaking aortic aneurysm was found. The court stated that the diagnosis had been missed but also stated that the surgical intervention probably would not have helped. Thus the outcome was attributed to the unavoidable consequences of the disease, not the failure to make a diagnosis.

In general, as accident compensation in New Zealand and Sweden has matured, judges have been more willing to recognize that a person whose bodily condition is made worse by medical treatment has suffered an injury caused by medical management. This means that a patient may seek compensation for an injury resulting from medical management even when that injury is a recognized risk of a treatment or procedure. The degree of risk must be considered, however. If a surgeon faces a patient with a leaking aortic aneurysm and the risk of death even with surgery is 75% while the risk of death without surgery is 100%, then death after attempted aneurysm resection is not compensable. On the other hand, when the risk of a poor outcome is in the 2-3% range and a poor outcome occurs, compensation is available. The threshold of risk below which one might compensate is not clear. Thus, while the trend[12] is toward broader interpretations of

---

[10] CM Oldertz, The Swedish Patient Insurance System, in Unexpected Complications in Medical Care, Stockholm, Almqvist and Wiksell 1979 at 244

[11] 4 NZAR 335 (1984)

[12] Gellhorn, Medical Malpractice Litigation (U.S.) -Medical Mishap Compensation (N.Z.), 73 Cornell L. Rev. 170 (1988)

causation, generic problems persist.

In short, determinations are never made with complete confidence. Rather than an either/or decision, the causation issue is often a continuum: some injuries clearly arising from medical management, others probably arising from medical management, and still others more likely arising from the disease process. All of the problems with causation noted above introduce a variety of shades of grey into an area that our tort system treats as black or white under the rubric of more probable or not determinations. In light of this, instead of asking our reviewers to make a determination regarding the presence or absence of causation, we asked them to grade their confidence in their judgment of the finding of medical management causation.

We used a six point scale for this grading of confidence in causation:

1) little or no evidence for causation
2) slight or modest evidence
3) causation not quite likely; less than 50/50, but close call
4) causation more likely than not; more than 50/50, but close call
5) strong evidence for causation
6) virtual certain evidence for causation.

Using this scale, we are able to identify the cases that are close calls for causation and others which are virtually certain for causation. This designation not only allows us to define a spectrum of causal determinations, but also allows us to determine from the entire population of cases the distribution of causal judgments between the comparatively easy and the comparatively hard cases.

The grading of confidence in adverse event judgments is critical to a consideration of how a no-fault system might function. Thus, we asked our reviewers to make definitive

decisions regarding adverse events: we did not permit them to make an ambiguous reply by choosing the midpoint (3.5) of the scale as a score. Because we had two physician reviews of each case, however, we could get average scores of 3.5. As a result, we defined adverse events as those cases with averaged scores greater than 3.5. Cases with averaged scores greater than 1.0, up to and including 3.5, were called low-threshold adverse events because we would include them as adverse events if our threshold were arbitrarily set lower. We called adverse events with scores of 3.0-4.0 close-call adverse events, because of their proximity to the midpoint of the adverse event causation scale. This information, along with the absolute numbers of adverse events, is crucial to the policy question of whether one could materially reduce the level of conflict and litigation about medical accidents by moving from a fault-based to a cause-based legal regime.

B. Judgments of Negligence

In our malpractice system, the patient's injury must have occurred through fault on the part of the doctor in order to be compensable.[13] Fault in malpractice cases means negligence, a failure on the part of the doctor to provide reasonably careful treatment. A great deal of legal experience with this standard makes it clear that the determination of malpractice almost always rests on a medical judgment. Did the treatment fall below the standard that normally should be expected from the

---

13 As mentioned, the medical compensation systems in New Zealand and Sweden do not make a determination of negligence but instead compensate those who have been injured whether or not their injury was caused by negligence. Since opposing sides do not need to argue over the presence of negligence, the injured party is compensated as soon as the evidence is clear on the nature and extent of the injury. Perhaps more importantly, since this approach separates the issue of negligence from that of compensation and since physicians are not individually liable, physicians are more willing to participate and, indeed, assist accident victims in seeking compensation. Unlike in the United States, where compensation for an accident signals negligence, physicians are more likely to play a pro-active role in the compensation process. This approach also encourages physicians and administrators to develop detailed data sets concerning injuries and thus an opportunity to develop systems for addressing injury prevention.

practitioner usually caring for this kind of disease? Nonetheless, many difficult conceptual issues emerge in making determinations of negligence that are distinct from those arising when making causal judgments.

First, many possible candidates may be used for the standard of care. Certainly, the standard of care in rural New York is different from that in urban teaching centers in the state. For example, in a teaching hospital in New York, a patient with angina refractory to medical therapy would probably immediately undergo cardiac catheterizations and be considered for surgical bypass. In rural New York, or in smaller community hospitals in large cities, these options might not be available or would be available only after risky transfer of the patient. The risks of transfer must be weighed against the risks of delay. Thus, the standard of care with regard to the timing of cardiac catheterization may be different from one hospital to the next. The law has in the past used a so-called locality rule to help make these determinations[14] although this rule has now largely been abandoned. Nonetheless, the problem persists.

We asked our physician-reviewers to use a statewide standard but to tailor their judgments to the particular circumstances of the hospital in question in 1984. This requirement is quite similar to the standard employed by Peer Review Organizations (PROs) evaluating care of Medicare patients across the country. We did not adopt a locality rule, and we did not seek local advice about the standard of care.

Another difficulty with judgments of negligence is that many factors go into the final determination, and these are rarely made explicit. Indeed, episodes of negligence may vary in character a great deal, according to the factors involved. We broke down the concept of negligent treatment into a number of

[14] The locality rule held that doctors were to meet the standard of care of a given community or geographic area, rather than a statewide standard.

constituents that the reviewer was asked to consider in making a judgment: the challenge posed by the patient's condition; the doctor's level of awareness of the risk in the treatment being used; the gravity of the harm that could materialize from that procedure; the degree of professional agreement concerning appropriate therapy; and the likelihood that this mishap would be repeated in the future.

Even making explicit the factors contributing to judgments of negligence could not guarantee, however, that every judge would develop the same rating. Every physician-reviewer must be prepared to elaborate his or her view regarding the standard of care in a given case. Since such judgments can never be completely objective, slight variations occur in estimates of the standard of care. These variations, in turn, can lead one reviewer to term a case negligent while another might term it non-negligent. We lack explicit standards regarding quality of care and so must rely on implicit standards. Although we tried to focus all reviewers' attention on the same components of negligence, we could not overcome all variation in assessments.

Consider the following examples. A surgeon is removing an injured spleen. The patient has had no previous surgery, and the spleen injury is uncomplicated. Nonetheless, the surgeon inadvertently cuts into the bowel, and bacterial contents from the bowel spill into the abdominal cavity. The patient develops fevers after the operation and requires intravenous antibiotics, prolonging the hospital stay. We would consider this an adverse event because the medical management -- the bowel perforation and the fevers resulting from the infection caused by abdominal cavity contamination -- prolonged the hospitalization. Many would say that the surgeon's inadvertent entry into the bowel failed to meet the standard of care and thus was negligent. In truth, this kind of complication can happen to even the best surgeon. Such an event raises quality of care concerns only if the problem happens with frequency in patients treated by this

surgeon. Nonetheless, in this episode, most would say the care was substandard.

Now consider a second case. The surgeon is again to remove a spleen, this time because the patient has a low platelet count. This patient also suffers from cancer of the lymph glands, or lymphoma, and he had had abdominal surgery previously. His lymphoma has been treated with radiation, and this therapy, as well as the previous surgery, has caused a great deal of scarring. Therefore, this surgeon, instead of encountering readily mobile bowel like that of the previous patient, comes upon a mass of scar tissue when he makes his first incision. He can barely recognize the spleen. He too enters the bowel, and again infection occurs.

We would term the increased length of stay necessary to treat the infection an adverse event, but we would likely not consider this adverse event negligent. Given the horrible state of the abdominal cavity, the standard of care is different than in the previous case. The background risk of any surgeon entering the bowel is much higher in this case than the previous one, and so we might not term this adverse event negligent. But the judgment depends not on some explicit set of criteria, but on the reviewer's judgement regarding the standard of care in this case (implicit criteria). Explicit criteria do not exist for every possible variation of care that may lead to adverse events.

Many other circumstances that could accompany surgical perforation of the bowel lie intermediate between these two examples. Some cases would be very difficult to call; one reviewer might find negligence, another might not. (Remember that every malpractice case that goes to trial features one expert who says the standard of care was met and another who says it was not.) Thus, we would expect greater unreliability in judgments of negligence.

We have used the same six point scale to judge negligence that we used to judge causation and adverse events (score > 3.5). Those adverse events for which the average negligence score was greater than 3.5 were labelled as negligent. This means of course that those adverse events which we term as negligent are not the only adverse events with some evidence of negligence. Those that we have so labelled meet the conventions described above. But adverse events are more accurately characterized as representing a spectrum from those with no evidence of negligence, to those with a great deal of evidence. The six point scale allowed us to determine the number of cases that were close calls for negligence, as well as those that were unequivocally negligent. We explore this notion further in later chapters.

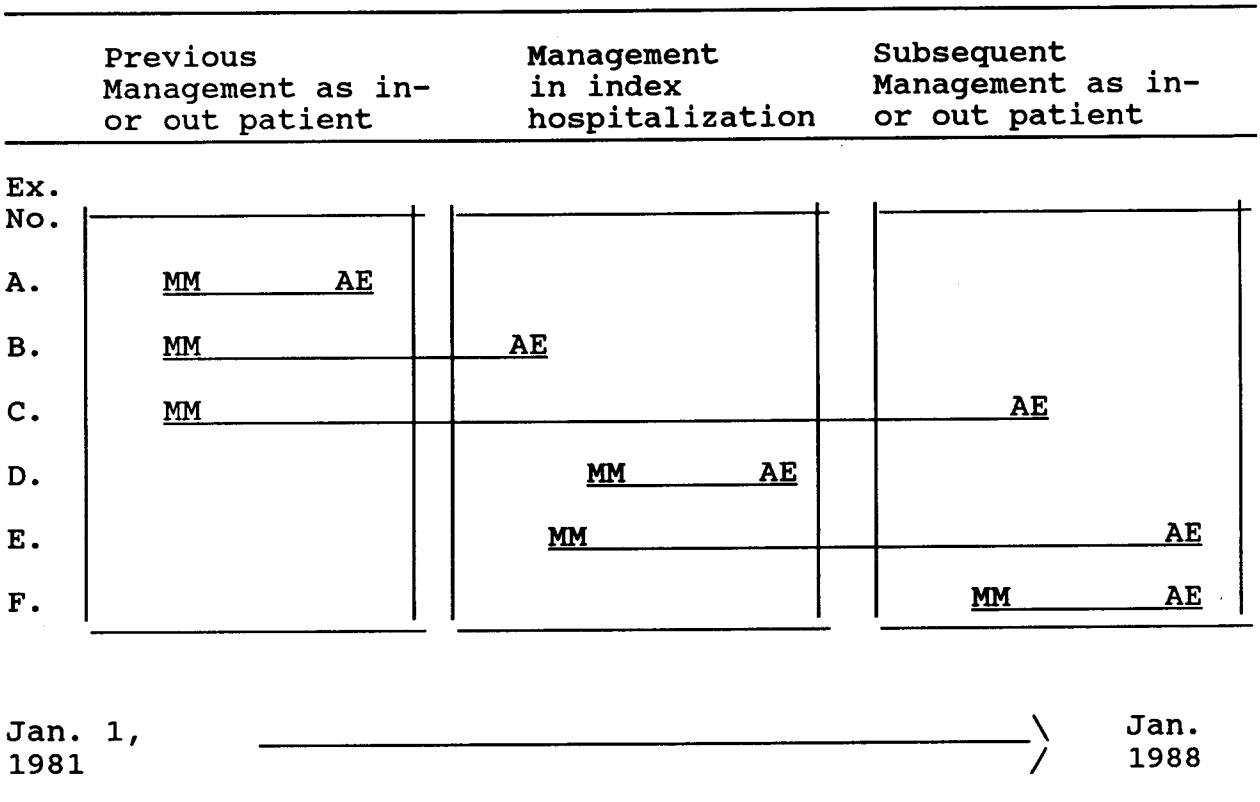
C. Defining Incidence

The previous discussion concerns the conceptual basis for judging adverse events and negligence. To arrive at estimates of incidence, however, we also had to consider the timing of the event. Our sampling strategy identified over 31,000 index hospitalizations in New York in 1984. Some patients may have had additional hospitalizations, information about which would appear in the medical record. In addition, some information about outpatient management was present in medical records. We wanted our reviewers to focus primarily on the index hospitalizations, but we were also concerned not to overlook adverse events occurring in previous or subsequent medical management. Further, we wanted to avoid overcounting adverse events.

For these reasons we sought to identify the several possible temporal relationships by which medical management and adverse events were linked to the index hospitalizations. We asked reviewers to identify adverse events that a) were a consequence of medical management prior to the index hospitalization and were discovered during the index hospitalization; b) occurred during

the index hospitalization and were discovered during the index hospitalization; and c) occurred during the index hospitalization but were discovered after discharge. Table 5.2 displays these relationships.

Table 5.2
TIMING OF MEDICAL MANAGEMENT AND ADVERSE EVENT



MM = Medical management causing adverse event (AE)

AE = Initial discovery of adverse event (not previously capable of being discovered by a knowledgeable patient or physician)

The possible combinations of MM and AE occurring in three time periods are detailed above. The middle block represents the information available from the record of the index hospitalization. The adverse events included in our Study in

principle are those labelled B, D and E, but our procedure precludes our identifying all cases of type E; hence, we included type B to compensate. To avoid the possibility of discovering the adverse results of a single course of medical management in every subsequent hospitalization, we instructed reviewers not to record adverse events discovered in a previous hospitalization. (See Technical Appendix 5.III.1.) These issues are re-visited in Chapter 6, wherein incidence categories 2 and 3 correspond to category B here, incidence category 1 corresponds to category D here, and incidence categories 4 and 5 correspond to incidence category E here.

In analyzing our data, we used various subsets of these incidence categories. In addition, we distinguished adverse events that resulted from previous or subsequent medical management in hospitals from those that occurred in an outpatient setting. This allowed us to develop some information on outpatient management of adverse events.

IV. Reliability and Validity Testing

Judgments at the screening phase are useful only insofar as they are valid, and judgments of causation and negligence by physicians are valuable only insofar as they are reliable and valid. Therefore, we undertook several studies, both in an intermediate pilot phase and during the review of records in New York, to reveal the validity and reliability of our processes for identifying adverse events.

Validity refers to accuracy, or the extent to which a measurement by a rater approximates the true value. The best test for validity would compare the results of a measurement process to a gold standard. Such a gold standard for iatrogenic injury is generally unavailable for implicit judgments of complicated clinical issues such as causation and negligence. For the MRA screens, the first stage of our process, however, we considered checks by a more senior MRA or a physician to approach

gold standard status, and thus we can focus on validity when evaluating the screening process.

Since no similar gold standard is available for physicians' judgments on causation and negligence, we must rely on other ways to determine validity. The content validity of a process is evaluated by asking experts to examine it and to comment on its appropriateness. Construct validity is assessed by comparing one measurement process to another.

We assessed the construct validity of our process by comparing results of the review process with a different method of review for the same records (part of our Intermediate Pilot Studies). We assessed content validity by comparing our record review results with results obtained in a review of litigation records. We describe these validity tests in detail below.

Reliability refers to the consistency of ratings or to the ability of various raters to reach the same conclusion about a specific case.[15] Previously reported studies concerning physician or medical-legal expert reviews of charts to identify adverse events have not addressed the issue of reliability.[16] Other physicians reviewing these same charts might reach entirely different conclusions than those reported. To evaluate the reliability of our judgments of causation and negligence, we undertook both an Intermediate Pilot Study of reliability and a reliability study in New York conducted concurrently with our main New York record review.

A. Validity of Screening Process

The first level review relied heavily on standard screening criteria. These screening criteria, which are clinical, were found to be highly associated with adverse events. The Medical

---

[15] J.C. Nummally, Psychometric theory, New York, McGraw Hill, 1978.

[16] For review of this literature, see Brennan, Localio & Laird, Reliability and Validity of Judgments Concerning Adverse Events, Med. Care in press.

Insurance Feasibility Study employed a set of 20 criteria. Dr. Mills developed statistics for the utility of each criterion and was kind enough to share this information with us. He found four criteria of little help in his study for the screening procedure because their independent yield was very low or absent and because legitimate hospital practices rather than potential adverse events were identified by these criteria.

Using his data, we developed our own set of 18 criteria for the Pilot Study in New York in 1986 to 1987. The criteria include previous hospitalization within six months; hospitalization within six months after the index hospitalization; transfer to intensive care while in hospital; transfer to another acute care facility; acute myocardial infarction following a surgical procedure; and neurological deficit at the time of discharge. The complete list of Mills' criteria is reproduced in Appendix 5B. Our own list of criteria for our Pilot Study are also in Appendix 5B.

Based on results of the Pilot Study conducted in New York in 1986 to 1987, we further revised the record screen and the instructions for its use. We decided that first-stage reviewers needed to identify only 1 of 17 possible criteria to refer cases for the second-stage review process. An 18th criterion, hospitalization longer than a diagnosis-specific threshold, would be identified by computer. These modifications decreased the time and, therefore, the cost for the first-stage review. As in the Pilot Study, all record screeners used a three-page instrument preprinted with selected data from the SPARCS data base to increase their productivity and accuracy. The final list of criteria that we used for the Medical Practice Study is given in Appendix 5B.

Selection and training of the MRAs were critical to success of the Study. To train them for both our Pilot and main Studies, we developed a manual describing the screening criteria and the

role of screening in the overall study design.[17] We introduced MRAs to this manual and the Study itself at a two-hour training program. Any questions that occurred to MRAs during the Study itself were referred to the MRA supervisor, who was an administrator recruited by HANYS (Hospital Association of New York State) and who oversaw the record review at a particular hospital.

Before initiating the record review, we designed and conducted an Intermediate Pilot Study to assess the ability of the MRAs, who were first-level record reviewers, to identify screening criteria in a valid fashion. Using a balanced incomplete block design, we assigned 360 medical records (including a subset known to have risk management problems) from two northeastern United States teaching hospitals to be screened independently by two MRAs. All 360 records were then reviewed independently by physicians associated with the Study to provide a standard for testing the sensitivity of the MRA-screening process.

We knew the importance of sensitivity for the initial screening review by MRAs: that is, to have a low false-negative rate. This study revealed a sensitivity for the MRA process of 84.5% (403/477) when checked by senior physician-screening reviews. When we evaluated the screening process in light of the final judgments of causation and negligence by senior physicians, the false-negative rate of the screening process for uncovering adverse events was 7.9% (19/241) and for negligence was 6.3% (9/144). (See Technical Appendix 5.IV.1.) The estimate of the negative predictive value of the screening was 99.5%.

In New York, we repeated this analysis in part. A random 1% of all records were independently screened by the MRA supervisor for presence of any of the 18 screening criteria. Overall, 314

[17] Hospital Record Screening Manual for Medical Practice Study, Appendix 5C.

records were in the 1% duplicate screen sample. Some records were missing at the first review; others were screened but lacked the duplicate screen. Among those missing a duplicate screen, we obtained an MRA review during the follow-up if the record had become available. In all, 282 records received a second, independent screen by the MRA supervisor. The results are presented in Table 5.3

Table 5.3
Comparison of MRA Screen and Supervisor Screen
in One Percent Sample

		Supervisor Screen: No Criteria	Supervisor Screen: Criteria Met	Total
MRA Screen	No criteria	191	9	200
	Criteria Met	9	73	82
	Total	200	82	282

Among the 82 records that were positive on the duplicate screen by the MRA supervisor, 9 were found to be negative on the initial screen. This represents a false-negative rate of 11% or a sensitivity of 89.0%. The corresponding sensitivity from the Intermediate Pilot Study was 84.5% (403/477). These figures from our 1% sample demonstrate even better validity than do the results of the intermediate study.[18]

B. Reliability of Physician Judgments of Causation and Negligence

We were quite concerned about the ability of our reviewers

[18] In the next chapter, we adjust our estimates of rates of adverse events and negligent adverse events using data from the Intermediate Pilot Study. We use Intermediate Pilot Study data because false-negative records in the 1% study in New York were not subject to physician review to determine whether they contained adverse events.

to render judgments on causation and negligence that were both reliable and valid. Therefore, we addressed issues of reliability and validity of judgments in the second level review process by physicians both in intermediate pilot studies and concurrent with our main study in New York. In our first Intermediate Pilot Study, we assessed the reliability and validity of physician ratings of causation and negligence in records from two teaching hospitals in the northeastern United States.

Using the Adverse Event Analysis Form (AEAF), ten physicians independently reviewed 225 records previously screened by medical record professionals. Each record was reviewed by four physicians in a balanced incomplete block design. A single reviewer was selectively unreliable, but averages of the scores of two independent physician-reviewers showed that the Study achieved good-to-excellent reliability in judgments of causation. In particular, we found the reliability (R_m) of the average of two physician-reviewer scores was .78 for causation, well within the excellent range. The reliability for averaged scores from two reviewers for negligence was .63, lower but still within the fair to good range. These results confirm our expectations that judgments of negligence have more reviewer variation than judgments of causation (see Technical Appendix 5.IV.2). They also indicated to us that each record would need to be reviewed by two physicians in the main study in New York. (One possible criticism of the California Medical Association study was that it was based on decisions by single reviewers. Possibly other physicians might not have arrived at the same conclusions on individual cases.)

In addition to the Intermediate Pilot Study of two teaching hospitals, we tested the reliability of our physician determination of adverse events and negligence in New York. All records with positive screens in two hospitals were subjected to the physician-review process on two separate occasions. On each

occasion the same MRA supervisor organized the screening process, but a different team of physician-reviewers carried out the reviews, a different MRA supervisor analyzed these reviews for potential discrepancies, and a different physician-supervisor was used to resolve any disagreements. A total of 318 records were screened as positive in these two hospitals. Table 5.4 shows the determination of adverse event and negligence for the two duplicate reviews.

Table 5.4
Results of Duplicate Physician Review

		Adverse Event					Negligence			
		Review 2					Review 2			
		Absent	Present				Absent	Present		
Review 1	Absent	249	21	270	Review 1	Absent	293	12	305	
	Present	13	35	48		Present	9	4	13	
		262	56	318			302	16	318	

For determination of adverse event status, there was agreement in 89.3% ([249+35]/318) of cases. The Kappa statistic gives a chance-corrected agreement of 61% (SE = ± 6%). The percentage of agreement for negligence is 93.4%; the Kappa statistic is 24% (± 11%). If we score a case as negligent when the combined confidence in negligence is greater than one (that is, if we include low-threshold negligence cases), the Kappa statistic increases to 47% (± 8%). Using the confidence of causation and negligence scores as ordinal measures of confidence, we compute the intraclass correlation coefficients as .70 for adverse events and .41 for negligence.

Overall, the New York study and the Intermediate Pilot Study had good agreement on causation judgments. They are made with

good reliability.[19] The results concerning negligence from New York demonstrate less reliability than our previous estimates although this score is similar to that encountered by other researchers attempting to make complex judgments regarding negligence.[20] Moreover, we expected the Kappa to be less in New York than in our intermediate pilot because the incidence of adverse events and negligence was lower in the New York reliability study.

In summary, probably because of the conceptual difficulties in defining negligence, our physician-reviewers were far from perfectly reliable when judging negligence. Our empirical data show that these judgments are even more difficult than judgments of causation. It is also interesting to note that if we base our analyses of reliability on all records with some evidence of negligence (that is, including low-threshold negligent adverse events), our reliability improves substantially. Thus, adverse events include a spectrum ranging from those with no evidence of negligence to those with a great deal of negligence. (This could be one explanation for a deterrent effect of the tort system on some adverse events that we classified as non-negligent.)

The lack of reliability regarding judgments of negligence does not in principle affect our estimate of the percentage of adverse events that are negligent. Reliability is like random error in any one case, and over a large number of cases, the error would average out. At the same time, lack of reliability does make it more difficult to show significant differences in the percentage of adverse events due to negligence rendered in subcategories, especially if those categories are small. That

---

[19] The Spearman Brown statistic used in the Intermediate Pilot Study was selected because we wanted to test the incremental value of adding more reviews to the average. The Kappa statistic was used in the New York study because we knew we would only be comparing the results of two review processes. Nonetheless, in large samples these statistics are broadly commensurate in terms of estimating agreement.

[20] Duboi, Rogers, Moxley, Draper & Brook, Hospital Inpatient Mortality: Is It a Predictor of Quality, 317 N. Engl. J. Med. 1674 (1987)

such differences do emerge, as shown in the next chapter, suggests real differences in rates of adverse events and negligent adverse events.

C. Validity of Physician Judgments of Causation and Negligence

1. Access to Specialists. In the Intermediate Pilot Study, we compared the review of medical records by board-eligible internists using our methods to a review undertaken by a group of internists and surgeons working together and advised by a larger group of subspecialists. We found that board-eligible internists without access to the opinions of subspecialists agreed to a large extent with board-certified internists and surgeons who did have such access (Kappa = .57 for causation, Kappa = .51 for negligence) (see Technical Appendix 5.IV.1). This suggests that the need for access to subspecialty advice in judgments of causation and negligence would have only modest incremental value. Thus, although our reviewers had access to experts in four subspecialty fields in New York, we believe that, with few exceptions, subspecialists were not needed to review particular subspecialty cases. This important circumstance, we believe, is a consequence of the nature of adverse events (they are comparatively obvious) and the determination of negligence (difficult for specialists and nonspecialists alike).

2. Completeness of Medical Record. We also needed to establish the validity of medical records as an adequate source of information for evaluating the quality of medical care. Lawyers, for instance, might argue that information from depositions and other forms of discovery are needed in order to uncover adverse events and negligence.

To address this question in the Intermediate Pilot Study, we compared the results of a review of the records in two northeastern United States hospitals with the results of a risk

management malpractice claims review. For each of the hospitals in this study, we checked for any pending or anticipated litigation concerning cases in the medical record review and we examined all malpractice claims filed up to a given date.

Our review of litigation records uncovered adverse events independently from the record review and thus provided both a check on the review process and a description of the malpractice claim that our methods did not identify. We found that physicians reviewing records identified 93 out of 116 of the adverse events eventually uncovered after lengthy risk management assessment of all records and interviews of providers or after the litigation process had uncovered information. This means the sensitivity of medical records for uncovering adverse events was 80%. (See Technical Appendix 5.IV.2.) For negligence, the sensitivity was slightly lower (76%). Kappa statistics indicate good agreement between the reviews of medical records and the reviews of the more comprehensive litigation records (causation Kappa = .57, negligence Kappa = .62). Moreover, the more severe adverse events were generally readily uncovered in a record review. Thus, we were satisfied that the judgments of physicians would be both reliable and valid.

V. Data Analysis

In this section, we describe in detail the methods we used to process and analyze our data from the main New York record review as well as the methods for the missing record follow-up.

A. Data Processing

Data processing began in the field at 51 survey hospitals. Overall instructions for the field review were contained in a "Guide for Directing the Medical Record Review," a document used by each team leader, a senior MRA supervisor in the hospitals. Our principal record tracking tool, a Leader Worksheet,

documented the identity of the hospital discharges sampled, our requests to the hospital for its production, and the record of reviews by MRA, physician, and senior physician. The hospital team leader completed this Worksheet, which later became a checklist for the Study's clinical data coordinator.

Each hospital received a list of the hospital records to be sampled two weeks before the first visit of the team leader to the medical record room. Records initially missing were then re-requested during the two- to four-week review period. Thus, hospitals usually had at least one month in which to find records in the sample.

The team leader was responsible for managing the initial screening process, notifying and assigning physicians to review referred charts, reviewing selected questions for discrepancies between physicians, and notifying the physician-supervisor of any discrepancies. At all times, the leaders were required to keep abstracting forms in sealed containers segregated from the remaining hospitals charts. At the conclusion of the hospital review, a Study employee personally carried all forms to Boston.

Upon receipt of the materials from each hospital, the clinical data coordinator in Boston reviewed each MRA record screen and Adverse Event Analysis Form to verify data. Using a pencil, the coordinator marked each response that was inconsistent either with other responses in the form or with explicit instructions in the "Physician Record Review Manual." These manual edit checks ensured that ambiguities in the review would not be lost or misinterpreted during computer analysis. The original responses were not obliterated, however, in order to facilitate subsequent error analyses. An in-depth description of these checks appears in Technical Appendix 5.V.1.

Upon completion of this initial manual check, all Hospital Record Screens and Adverse Event Analysis Forms were keypunched and verified. Using computer data bases, we then merged data

from the two physician reviews and any senior physician review to check further the validity and consistency of responses.

Errors could be divided largely into two groups: missing reviews and incomplete reviews (missing items). Among the cases with two physician reviews present, 48 (0.6%) lacked a senior physician review to resolve a disagreement on the choice of the adverse event. For these we randomly selected one of the two reviews to use in the analysis. For 41 cases in which one or more data items on a review were missing, we used imputation to complete the missing data items conditional on a reviewer's response to other items (see Technical Appendix 5.V.1).

During all phases of data processing and storage, both manual and computerized, we maintained security and confidentiality by using internal, randomly assigned case identifiers rather than hospital name, patient name, or medical record number.

B. Variable Definitions

In the results reported here, a single determination of adverse event and negligence status was made from the average scores of the two reviews for a case. We use the following cut points: an adverse event requires an average confidence of causation score of 4 or higher; a negligent adverse event is an adverse event with a confidence of negligence score of 4 or higher. A causation score greater than 1, but less than 4, is a low-threshold adverse event; a causation score of 3, 3.5 or 4 is a close call.

For cases with only one physician review (those for which the supervisor found it necessary to override both the conflicting reviews, and a few cases for which only one review was obtained), the determination was made using the single review. The results for disability are based on the ratings of both reviewers (except in the case of single reviews) by

assigning one half of the weight for each case to each of the two reviewers.

As discussed in previous sections, we had reviewers rate both the events that occurred in the sampled (index) hospitalization (and were possibly discovered subsequently) and those that occurred prior to the index hospitalization. In presenting the results of the record review, we use different definitions of adverse events according to the purpose of the analysis.

For the purpose of developing an annual rate of injuries, we considered only those adverse events that were first discovered in the 1984 index hospitalization, regardless of when or where they occurred. This method of counting adverse events (as opposed to counting only 1984 occurrences) allows us to capture adverse events that occur in any patient/provider encounter but that are not discovered until a subsequent hospitalization. If we were to count all adverse events that either occurred or were first discovered in the sampled hospitalization, we would overestimate the annual rate of events (see Technical Appendix 5.III.1). Demographic information on patients was obtained from the SPARCS data file as described in Chapter 4.

We used two different methods to determine disability scores. Each adverse event was reviewed by two reviewers. While in nearly two thirds of the cases the reviewers agreed, in the remaining one third of cases there were different disability scores.

For the population estimate we assigned one half of the case weight to each review and then summed. This method has the advantage of capturing the information for every review, but it does not allow us to assign a disability to each case.

Since we wanted to complete analyses on specified subsets of cases, we developed another method to determine disability for each case. We randomly selected one reviewer's disability score when reviewers disagreed. This method does not allow us to

capture information for each review but does allow us to assign a disability score to each case. In this latter method, we also eliminated the "not determined" category. The results of the two approaches are quite similar (see Technical Appendix 5.VI.1).

C. Adjustments of Sample Weights for Missing Records

A common problem in surveys is missing records. Although no substitute for obtaining all or nearly all records is ever thoroughly satisfactory, attaining 100% response from all record rooms in all 51 hospitals is clearly not possible. With the demands of patients returning to hospitals for follow-up care, utilization review, and microfiche copying for long-term storage, a medical record room rarely has 100% of its files available. Because we drew a random sample of records within hospitals, we faced a locate rate of less than 100%.

An analysis based on located records alone will be biased if the non-located records differ systematically from the located and reviewed charts. A standard method of dealing with non-response in sample surveys is to adjust (within subclasses) the sampling weights of the reviewed records upward to account for the weights of the missing records. This method is effective in eliminating bias only if the adverse event rate among the located records is the same as the adverse event rate among the missing records within adjustment subclasses. If this condition is true, the non-locate problem is said to be ignorable,[21] and the analysis based on the located records using adjustment weights gives unbiased results.

Rather than rely on this crucial assumption, we chose to undertake a follow-up survey. In a random selection of six hospitals we sought to determine whether the rate of adverse events differed among records reviewed at the initial visit and records found and reviewed at follow-up. The missing record

---

[21] R.J.A. Little & D.B. Rubin, Statistical Analysis with Missing Data, New York, Wiley, 9-10, 1987

follow-up is discussed and its technical details are presented in Section V.E of this chapter. In this section we describe briefly the adjustment to the sample weights to account for records not located at the initial visit. Technical Appendix 5.V.2 outlines additional details of our methods.

1. Records Not Located. Although the MRAs were able to locate over 96% of the records on our list during our initial visits to the hospitals, the rate of locating records varied by institution. In five of 51 hospitals, we found 100% of records. In contrast, at one facility we could not gain access to over 20% of the files during the initial visits because of difficulties with warehouse storage. Table 5.5 lists the locate rates for all hospitals in the sample; actual numbers are omitted to protect the identity of individual institutions. Overall, 3.9% of the records were not located during the first review.

In adjusting the weights of the located records to account for the non-located, we had to rely exclusively on SPARCS, the only available data on these records. Our goal was to adjust the located records by a method that would take into account differences in adverse event rates among different groups of patients as defined by available SPARCS data.

Table 5.5
Records Missing After Initial Hospital Visits

Hosp No.	Fraction Missing
1	0.111
2	0.042
3	0.071
4	0.006
5	0.021
6	0.028
7	0.024
8	0.006
9	0.002
10	0.038
11	0.107
12	0.002
13	0.006
14	0.020
15	0.056
16	0.026
17	0.008
18	0.048
19	0.068[1]
20	0.072
21	0.014
22	0.061
23	0.033
24	0.045
25	0.050
26	0.039
27	0.082
28	0.049
29	0.204
30	0.004
31	0.136
32	0.078
33	0.025
34	0.014
35	0.046
36	0.019
37	0.185
38	0.011
39	0.030
40	0.013
41	0.021
42	0.020
43	0.040
44	0.012
45	0.008
Hospitals With All Records Located	
46	0.000
47	0.000
48	0.000
49	0.000
50	0.000

	Sample	Missing	%
TOTALS	31429	1234	0.039

[1] This entry accounts for two hospital divisions that shared a single record room.

2. Records Located But Not Reviewed. A second missing record issue arose once the medical records were located and screened for 18 criteria. Of the 30,195 records located for review, 7,817 passed the first stage of screening. For a variety of reasons, 74 of these records were not reviewed by physicians. Some originally located by the MRAs were no longer available when the physicians came to the hospital. Others were inadvertently misdirected by the MRA and never received a second stage review. In any event, the records were not available for review when the team of physicians arrived. Such records referred by MRAs but not reviewed by physicians needed a separate adjustment because, not having passed the initial screen, they had a much higher probability of including an adverse event. Our adjustment therefore had to account for the increased likelihood that these 74 cases would, if reviewed, have been classified as having adverse events.

3. Methods for Adjustment. We adjusted for both the missing records and those referred but not reviewed by using a common procedure called raking ratio estimation.[22] This procedure, described in more detail in Technical Appendix 5.V.2, assigns the sampling weights of the missing records to those actually located and then assigns the weights of the cases not reviewed to those with completed reviews. By first dividing the records into adjustment classes formed by hospital and four weighting groups within hospital (elderly, oversampled high-risk diagnoses, low-risk births, and all others), we took into account different adverse event rates and locate rates within these classes.

---

[22] Oh & Scheuren, Weighting adjustment for unit nonresponse. In Incomplete Data in Sample Surveys: Theory and Bibliographies. New York: Academic Press, 143-67 (W.G. Madow, I. Olkin & D.B. Rubin eds. 1983)

D. Analytic Methods

Because we chose patients with probabilities depending upon age and DRG (as described in Chapter 4), each observation had to be weighted inversely by the probability of selection in order to obtain figures reflecting the statewide population. As described previously, the original sampling weights have been adjusted to account for missing records and records that were missing a physician review. In all analyses in Chapter 6, we use these adjusted weights to obtain population rates and projections unless explicitly stated otherwise.

Standard errors for rates and population projections take into account the clustered nature of the data (as described fully in Technical Appendix 5.V.3). We used the SESUDAAN software package produced by RTI[23] for computation of standard errors and t-tests. Wald tests, based on these standard errors, were used for comparisons involving more than two groups (Technical Appendix 5.V.3).

To compare adverse event rates for different subgroups, we used directly standardized rates, stratifying on patient age (using 5 categories displayed in Table 6.6) and DRG category. We used four DRG categories obtained as follows. Prior to gathering the data, we asked three physicians involved in design of the Study to rate all 470 DRGs on a scale from 1 to 6, reflecting their impression that the DRG on the basis of its complexity was more (6) or less (1) likely to be associated with an adverse event. We averaged the three ratings and made three cut points in the distribution of average rating to define four categories of DRG. (See Technical Appendix 5.V.4.)

Table 5.6 shows population estimates of adverse event and negligence rates by DRG category. There was a monotonic increase

---

[23] B.V. Shah, SESUDAAN: Standard Errors Program for Computing of Standardized Rates from Sample Survey Data. Research Triangle Park, N.C., Research Triangle Inst. (1981)

in adverse event rate with increasing DRG category. The differences were statistically significant. The percentage of adverse events that were negligent did not significantly differ by DRG category. This was expected for two reasons. First, reviewers were asked to include the factor of case complexity in their evaluation of negligence; and second, the rate of negligent adverse events computed as a percentage of all adverse events controls for complexity by conditioning on the number of adverse events in the category.

Table 5.6

Population Estimates of Adverse Event and Negligence Rates by DRG Category

DRG	Adverse Events		Negligent Adverse Events	
Category	%	Standard error	%	Standard error
1	1.82	.32	24.4	6.6
2	3.34	.21	30.0	3.4
3	4.26	.36	27.0	2.9
4	7.13	.91	23.0	4.6
p value[1]	p < .0001		p = .47	

1 Wald Statistic

We also present the percentage of adverse events that are due to negligence by subgroups. In comparing these rates by subgroup, we standardized for patient age only, using only two strata (less than 65, 65 or over). The relatively small numbers of adverse events made it infeasible to stratify more finely and, more importantly, the overall rates of percentage of adverse events due to negligence show no association with DRG level (Table 5.6) or with age, except for a higher rate among patients 65 or over. Details of the standardization procedure are given in Technical Appendix 5.V.5.

E. Missing Record Follow-Up

The primary results presented in Chapter 6 are based on records reviewed during the initial visit to a hospital. We used adjusted weights to account for records missed on first review, but the validity of our missing data adjustment process relies on a crucial assumption about the missing records: within hospital and weighting class, the rate of adverse events is the same among the reviewed and missing records. The missing record follow-up was designed to test that assumption and to provide data for adjusting our final figures to account for any non-locate bias.

Our follow-up of missing records had three purposes:

(i) The SPARCS data do not contain patients' names, which are necessary to match a case in our sample with data on claims filed. One purpose of the follow-up, then, was to obtain patients' names for all missing records to facilitate the claims matching (see Chapter 8). To this end, we attempted to obtain the patient name and Social Security number on all records missing at the first review.

(ii) Because we needed information on compensation from claims paid to patients with adverse events, we were especially concerned about reviewing all records involving litigation for the presence of adverse events and negligence. Thus, a second purpose of the follow-up was to obtain physician review for any record having legal status (one of the previously discussed 18 screening criteria). All records missing in the first review that could be located were screened by MRAs for their legal status. If they were labeled legal, they were reviewed by a physician.

(iii) We wanted to determine whether the overall adverse event rate was different in records not initially located. Because a full review for all missing records in all hospitals

was infeasible, we used the double sampling approach.[24] We selected, at random, six hospitals for follow-up with probability proportional to the number of missing records. At these six hospitals we made intensive efforts to locate all records not found at initial visit; all located records were given the full two-stage review.

The missing record follow-up took place between four and six months after the initial visit.

In order to test for non-locate bias, we used only the data from the six randomly selected hospitals. We compared the weighted rates of adverse events in the records initially located at the six hospitals (using weights unadjusted for missing records) with the corresponding rates estimated from the records located on follow-up at those same six hospitals. In order to adjust for non-response bias, we use a weighted combination for the rate estimated from the follow-up data only and the rate estimated from the initial locates only. This is a version of the standard double sampling estimator proposed by Hansen and Hurwitz. (See Technical Appendix 5.V.6. for details.)

---

[24] Hansen & Hurwitz, The Problem of Nonresponse in Sample Surveys, 41 J. Am. Stat. Assoc., 517-29 (1946)

ADJUSTMENT FOR DOUBLE COUNTING OF ADVERSE EVENTS

One objective of the Medical Practice Study was to estimate the overall number of AEs in New York State in 1984. Another objective was to estimate the costs associated with these AEs. In addition to making these overall estimates, we wanted to make inter hospital comparisons based on the location of medical management that produced the AEs and to model the effects of litigation on the costs and quality of care. To make these estimates and comparisons using a single record review process and a single data base, we needed to adopt various conventions in counting AEs. This appendix outlines those conventions and the reasons for their adoption.

I. The Problem: The Index Hospitalization and the Continuum of Medical Care

Our medical record review protocols have assumed among other things: (a) that medical management often occurs at one time and the resulting AE manifests itself later; (b) that medical management in the physician's office can produce AEs, that can render the outpatient sick enough to need hospitalization; and (c) that patients who suffer AEs might need to be admitted to another hospital for follow-up care at a later time. These assumptions reflect the fact that AEs do not occur instantaneously during a hospital stay; they begin with medical care and sometimes develop over time both inside and outside the hospital.

This delay between cause and discovery of the AE would present problems for estimating the incidence and variation of AEs even in a review of all 2.7 million records for 1984. Our sampling design reflects this inherent complication. To the extent that AEs arise out of medical management during one admission and result in a later readmission during which the AE

is first discovered, a single AE can be counted via more than one hospitalization. By their very nature AEs can span more than a single hospitalization.

Double counting of a single AE can occur in the following situation. If we sampled the hospitalization of medical management, a reviewer following the Study protocols could identify the cause of an AE, follow the hospital record until the re-hospitalization for its discovery, and determine that an AE occurred. Likewise, if we sampled the hospitalization in which the AE was discovered, the reviewer could look back in the medical record for the medical management in the prior hospitalization and again find the AE. Thus, the same AE could be discovered via more than one sampled hospitalization.

Similar problems occur regularly in sample surveys under the heading of duplicate listings. [1] A common solution is a counting rule that eliminates the possibility of overcounting. In our case, we could solve the problem if we counted either all AEs caused in 1984 regardless of the time of discovery or all AEs discovered in 1984 regardless of the time of cause. All AEs not fitting the counting rule would be eliminated.

II. Date of Medical Management and Date of Discovery of Adverse Event

The problem of possible overcounting of AEs discovered via a review of medical records is not new; it faced the California Medical Association Study in 1974. To resolve the problem and provide data for different aspects of the Study, we collected data on the timing of medical management in relation to discovery of an AE. In question 6.41 of the Adverse Event Analysis Form, we asked the reviewer to place cases into 5 categories:

---

[1] Kish L. Survey sampling. New York: Wiley, 1965:58, 388-396.

(i.) medical management occurred and the AE was discovered during the index hospitalization,
(ii.) medical management occurred during the index hospitalization, but the AE was not discovered until outpatient treatment (location not specified but presumed to be at the same hospital) after discharge,
(iii.) medical management occurred during the index hospitalization, but the AE was not discovered until a subsequent admission to the same hospital,
(iv.) medical management occurred during outpatient treatment (either at a hospital or at a physician's office) prior to the index hospitalization, and the AE was discovered during the index hospitalization,
(v.) medical management occurred during a prior admission (perhaps in a different hospital, but not certain) but was discovered during the index hospitalization.

We also provided a diagram to assist the reviewer in this analysis (see main text). If medical management occurred prior to the index hospitalization, or if the AE was first discovered after the index, we asked for the appropriate dates in questions 6.42 and 6.43.

III. Counting Rules Adopted

A. Calculating Incidence of Adverse Events

For determining the incidence of AEs, we chose the same counting convention as did the California Medical Association Study in 1974. Using the above-defined categories, we counted only AEs in categories i, iv, and v. The categories include only those AEs discovered in the sampled hospitalizations in 1984.

The advantage of using discoveries for an overall incidence rate lies in the ability of reviewers to identify AEs more readily by looking retrospectively from the hospitalization of discovery backward to the date of medical management, rather than by examining the medical management and trying to follow the patient forward until manifestation of the AE. The advantage stems chiefly from the fact of inter-hospital patient transfers and referrals. Where medical management in hospital #1 gives rise to an AE discovered in hospital #2, the AE can never be identified by a review only of hospitalizations of medical management. On the other hand, by examining and counting AEs from the perspective of the hospitalization of discovery (#2), the reviewer can usually determine from the patient history whether the care at hospital #1 caused the injury. For this reason we chose to adopt the rule of counting as AEs only those AEs discovered in the sampled hospitalization in 1984.

B. Hospital-specific Adverse Event Rates

For analyses comparing hospitals, our focus was the location of medical management rather than the place of discovery of the AE. A counting rule that assigned an AE to the hospital of discovery would result in a bias against referral centers, in which patients with serious complications are often treated. Only in AEs in category 1, 2 or 3 (see categories above), did the medical management occur with certainty in the sampled hospital.

Category 4 cases result from outpatient care and thus are not attributable to the sampled hospital. Category 5 AEs could result from medical management in another hospital. Future analyses will determine which of these AEs occurred in the sampled hospitals. For the study of deterrence, we shall also confine the AE categories to those known to have occurred in sample hospitals.

C. Adverse Event and Negligence Rates for the Study of Litigation

In Chapter 7, our measures of the incidence of litigation also depend on the appropriate definition of the incidence of AEs. To allow for comparisons of population estimates of AEs, negligence and litigation, we chose to limit our analysis and estimate of AEs in litigation to those in categories i, iv, and v.

D. Describing Characteristics of Adverse Events

For describing AE characteristics (Chapter 6), we also confined our analyses to AEs in categories i, iv, and v.

RELIABILITY AND VALIDITY OF PHYSICIANS' JUDGMENTS AND VALIDITY OF MEDICAL RECORD ROOM ADMINISTRATORS' SCREENS

I. Introduction

Previous efforts to estimate rates of hospital adverse events have relied largely on medical record reviews by physicians and medical-legal experts. This type of implicit review process, relying on physician judgments, has generally demonstrated little methodological rigor. If calculation of rates of iatrogenic injury and assessment of the related issue of quality of care are to be useful, the concepts of reliability and validity must be applied.

Reliability refers to the consistency of ratings or to the ability of various raters to reach the same conclusion about a specific case. Previously reported studies concerning physician or medical-legal expert reviews of charts to identify AEs have not addressed reliability. Other physicians reviewing these same charts might reach entirely different conclusions than those reported. Validity refers to accuracy, or to the extent to which a measurement by a rater approximates the true value. The best test for validity would compare the results of a measurement process to a so-called gold standard. Such a gold standard evaluation of iatrogenic injury is generally unavailable for implicit reviews, given the uncertainties involved in most medical interventions and the difficulties associated with development of clinical standards. Therefore, one must rely on other ways to determine validity.

The content validity of a process is evaluated by asking experts to examine the process and comment on its appropriateness. An example would be the evaluation of a hospital's quality assurance process by a Joint Commission for

Accreditation of Health Care Organizations (JCAHO) official. Face validity is elicited by asking the raters using a process whether it seems reasonable. An example of this would be querying physicians performing Peer Review Organization (PRO) reviews about the process they are using. Construct validity is assessed by comparing one measurement process to another. An example would be a comparison of the hospital quality assurance program with the PRO process. Physician review of medical records to identify AEs provides a great deal of face validity but little construct validity. Thus generalizations from reports based on physician reviews are compromised by the failure to assess reliability and the incomplete attention to validity.

For example, the Medical Insurance Feasibility Study, the only previous effort to estimate the adverse event rate in hospitalized patients, relied on a medical record administrator (MRA) screening process followed by a single medical-legal judgment to estimate the rate of adverse events in hospitals. The researchers performed a validity analysis on a subsample of the charts reviewed by MRAs and found that 12% of charts read as negative were positive for screening criteria. The only assessment of the quality of the medical-legal expert review was a re-review of a subsample of charts read as negative on the first expert review. Ten of 415 of these charts contained AEs. The researchers did not discuss how the subsamples were drawn, nor did they mention whether the second review was blinded. No formal reliability or validity testing of either the MRA screening review or the medical-legal expert review was reported.

The success of a population-based study of the incidence of **adverse events** depends critically on a method which produces valid and reliable judgments about **adverse events** and negligence. This appendix describes the results of a pilot study we designed

to assess the reliability and validity of our physician judgments and the validity of screening judgments by MRAs.

II. Methods

A. Sample Selection

The sample consisted of 360 medical records from two teaching hospitals in the Boston area. The hospitals were chosen for their close proximity to our study group and for the willingness of their risk managers to cooperate with the study. The Institutional Review Board of each institution approved the study. One hundred and eighty records were identified through risk management claim files. A claim file exists because: (a) the patient filed a malpractice suit; (b) the patient did not file a suit but sought compensation from the hospital or physicians; or (c) the hospital and risk management office considered a medical event to be serious enough to warrant an investigation. The hospitalization that gave rise to the claim was designated as the index hospitalization and the medical records on these were requested. The hospitalizations occurred from January, 1984 through May, 1987. These records were selected because they were thought to have higher incidence of adverse events than would randomly selected records.

A set of non-risk management charts was compiled by randomly selecting an equal number of hospitalizations from 1984-1987 from the two teaching hospitals. This process resulted in 108 risk management and 108 non-risk management records in hospital A, and 72 risk management and 72 non-risk management records in hospital B.

B. Format for Identification of Adverse Events

Our method for identifying **adverse events** is a modification of the two-stage process employed by Mills et al. in their Medical Insurance Feasibility Study. The first stage is a screening process involving 18 criteria that consist of readily identifiable events often associated with poor patient outcome. The criteria are retrospective (e.g., previous hospitalization within one year of index hospitalization), intercurrent (e.g., transfer to intensive care unit; length of stay outlier), discharge specific (e.g., discharged with new neurological deficit), or prospective (e.g., rehospitalized within six months of index hospitalization). This review process is completed by medical record room administrators (MRAs). MRAs are record room technicians who are familiar with medical records and are usually employed to code diagnoses.

The second stage is a more detailed analysis by physicians of those medical records that were positive for one of the screening criteria. Each record is examined using a structured adverse event analysis form (AEAF) that focuses the judgment of the reviewer regarding the cause of the patient's disability as well as the standard of care manifested by the medical record. The AEAF was developed using the multi-attribute utility process described by Gustafson. We convened a panel of experts in clinical decision-making and quality assurance, and they gave their opinions on the important attributes of **adverse events**. We then formulated these insights into specific criteria.

Using the AEAF, the reviewer looks for evidence of an injury in the form of a prolonged hospital stay or a disability at the time of discharge. The reviewer then must decide whether the injury was caused by medical management or by the disease

process. He or she indicates his or her judgement regarding causation on a continuous scale of 0 to 10 with 0 denoting no possibility of causation by medical management and 10 being unequivocal causation by medical management. Any intermediate score except for 5.0 is permitted. If the causation score exceeds 5.0, the case is considered to be an **adverse event.** The score of 5.0 is excluded because we want reviewers to answer the causation question unequivocaly, as a court of law must.

Any case with even minimal evidence of medical management causation is evaluated for negligence. Negligence is defined as failure to meet the standard of care of the average practitioner for the particular year in which the care was provided. The reviewer considers certain aggravating and mitigating circumstances before making this judgment. Again, the judgments are made on a continuous scale from 0 to 10, excluding the midpoint.

C. Design of Pilot Study

First, MRAs completed the initial screening. Each chart found positive for a screening criterion by an MRA was referred to the physician-reviewer stage for **adverse event** analysis. The 10 physician reviewers were senior residents or fellows recruited from training programs in the Harvard teaching hospitals. One of the investigators gave the physician-reviewers a two-hour training course in the use of the AEAF, and the reviewers studied a comprehensive 30-page manual describing the review process and the concept of an **adverse event.** One complete review of each record comprised the MRA screening stage and the physician-reviewer **adverse event** analysis stage.

Another complete review of each record, including both stages of the review process, was completed by senior physicians.

Two board-certified internists, three board-certified surgeons, and one board-eligible senior internist, all of whom were involved in design of the study. They did the initial screening using the same forms as the MRAs, and then, on those records with positive screening criteria, they did the **adverse event** analysis. The results of these two parallel and independent reviews provided the basis for the studies of validity and reliability (described below).

D. Medical Record Administrator Study

With the help of a 25-page manual explaining the rationale underlying the choice of screening criteria, an experienced physician trained MRAs in a three-hour session to use the Hospital Record Screening Form containing the sixteen criteria. Since MRAs are experienced in screening charts for diagnoses and procedures, it was very easy to train them to look for the criteria. An administrator who had helped design the screening criteria supervised the review process.

Six MRAs participated in the study, each one reviewing 120 hospital records, creating a total of 720 MRA reviews. We assigned the risk management and non-risk management records randomly to reviewers using a balanced incomplete block design (BIBD), which ensured that each pair of MRAs reviewed 24 charts in common. None of the reviewers knew the risk management/non-risk management status of the records. By this design, each record underwent two independent reviews by MRAs.

We determined the validity of the MRA review by comparing it with reviews by senior physicians. We treated the senior physician review as a gold standard and thus calculated the

sensitivity of the MRA process. We also estimated the prevalence of referred charts in a hospital population, using preliminary data from our review of 30,000 medical records in New York State, and the incidence of adverse events in a hospital population, using the Medical Insurance Feasibility Study, and then calculated the negative and positive predictive value of the screening process. To assess the errors that MRAs made in applying the screening criteria, we evaluated the individual error rates for each MRA. We then used a linear model to calculate the likelihood ratio test statistic on the null hypothesis that there was no variation from MRA to MRA.

E. Physician-Reviewer Reliability Study

We assigned the 225 records with positive screening criteria (161 risk management and 64 non-risk management) in groups of 15 to the 10 physician-reviewers in a balanced incomplete block design so that each physician would read 6 groups (90 records in all). In this way, each record received 4 independent physician reviews. Physician-reviewers were blind to the risk management status of the record. They were given a copy of the MRAs hospital record screen. Working independently, the physician-reviewers completed an AEAF for each case reviewed.

We used this data from repetitive reviews to calculate the reliability of physician-reviewer judgments regarding AEs. Reliability of the physician causation and negligence scores was measured using two scales: first, the 10-point scale of the AEAF, and second, a binary scale in which the 10-point scale score was characterized as above or below the midpoint. The intra-class correlation coefficient, R, was used to measure reliability in all cases. When considering two reviewers for each record, the intra-class correlation coefficient is quite similar to the ordinary correlation coefficient for pairs of

measurements. With a large number of records it is equivalent to the kappa measurement of agreement when using the binary scale. The intra-class coefficient has the advantage of being easily calculated with more than 2 but less than 10 reviewers per record. The R's were calculated using the SAS VARCOMP procedure, with maximum likelihood estimators, treating the reviewers as random effects.

To estimate the reliability of the average score of more than one reviewer, we used the Spearman-Brown formula for stepped-up reliability. The reliability of the mean of m independent replicate measurements is given by:

$$R_m = \frac{m^* \hat{R}}{1 + (m-1)\hat{R}}$$, where R is the reliability of a single measurement, as described above.

Although there is no universal standard of what value of R or R_m reflect reliable judgments, commonly recognized ranges are poor reliability for R less than 0.4, fair to good for R between 0.4 and 0.75, and excellent for values above 0.75.

The jackknife method provides an estimate for the variance of R and for approximate confidence limits for this complex statistic. We used it to evaluate further the reliability for cause on the 10-point scale. In addition, this method allowed us to calculate the impact of any one physician on overall reliability.

Our record selection process was deliberately designed to include more AEs than one would expect in a random sample of hospital discharges in order to permit the accuracy needed to determine the false-negative rate of the objective portion of the review process. Ideally, however, reliability should be tested

using a random sample of the target population because standard measures of reliability depend upon both the association between repeat reviews and the prevalence of the event in the sample. One measure of association between repeat reviews which does not depend upon prevalence is the odds ratio. In this setting, the odds ratio gives the odds favoring a positive second review, given that the first review is positive, relative to the odds favoring a positive second review given the first review is negative. Thus we adjusted for our sample selection and estimated the reliability in the target population by using the sample estimate of the odds between any pair of reviewers and by making assumptions about the **adverse event** rate.

F. Senior Physician Correlation Study

For all 360 medical records, the senior physicians completed the first step screening process, with 4 senior physicians doing 72 charts and two doing 36 charts each. If they found that no screening criteria were met, the review process stopped. If screening criteria were met, the chart was referred for second stage review. Any charts for which MRAs found positive criteria but the senior physicians found no criteria were re-reviewed by one the investigators, who was also a senior physician. If the MRA finding was verified, the chart was re-classified as positive for screening criteria.

The senior physicians completed the second stage review process using the AEAF. After completion of the screening process and the initial **adverse event** analysis, records were reassigned for a second **adverse event** analysis to a team consisting of two of the other five senior physicians. Working together and using the AEAF completed by the initial senior physician-reviewer, these two physicians performed a second **adverse event** analysis and jointly filled out another AEAF. If

they agreed with the initial review, they entered a final "consensus score" for causation and negligence. The senior physician review thus involved open discussion of each case. (In contrast, the physician-reviewer study was closed, with no discussion among the physician-reviewers.)

If the first and second senior physician-reviewers disagreed on significant issues for a case, or if policy questions arose, the case went to a conference. At the conference, consisting of at least 4 of the 6 senior physicians, there was open discussion on both policy issues and on the details of the individual cases and development of consensus scores. Where judgments were beyond the clinical expertise of the group, specialty consultations were obtained. The consultants included senior faculty in neurosurgery, obstetrics-gynecology, orthopedic surgery and neonatology at the Harvard Medical School. After consultation, the case returned for determination of a final consensus score by the conference session.

The results of this senior physician review were used in two ways. First, the information from the initial screening stage was used to validate results of the MRA study by acting as a gold standard for specificity and sensitivity calculations. Second, we used the consensus scores from the **adverse event** analysis by the senior physicians to assess the construct validity of the **adverse event** analyses completed on the 225 records by the physician-reviewers in the reliability study. To develop a single composite binary causation score for each record from the reliability study, we characterized the mean causation scores of the four physician-reviewers as being above (positive for causation) or below (negative for causation) the midpoint. For the senior physician reviews, we characterized the corresponding consensus score in a similar fashion. The sets of binary scores from the physician-reviewers and senior physicians were evaluated

for agreement using a kappa statistic. We followed the same process with regard to negligence scores.

In addition, we classified as difficult those cases that went to the senior physicians consensus sessions because primary and secondary reviews by senior physicians had resulted in causation or negligence scores on opposite sides of the midpoint. For the difficult cases that were also a part of the 225 records in the reliability study, we calculated separate scores for inter-rater agreement.

G. Confidentiality

MRAs, physician-reviewers and senior physicians received repeated instructions concerning the importance of confidentiality. A code number specific to this study was the sole identifier of all data collection forms. The risk management records, as well as the identification code, were locked at a central office. After completion of the data collection and analysis, we destroyed the medical record identifiers.

III. Results

A. Medical Record Administrator Study

MRAs performed 688 screening evaluations out of a possible 720 reviews (360 cases times 2 reviews). Thus 4.4% of reviews were missing. Since the senior physicians could not complete evaluation on several charts due to record room scheduling difficulties, the study generated 672 pairs of MRA and physician reviews. The sensitivity for MRA analysis was 84.5% (403/477). Specificity was 71.8%, (Table 1). Among the 57 cases (74 reviews) in which the MRA missed positive screening criteria,

both MRAs missed the criteria in 17 cases. These seventeen cases account for 46% of the missed reviews (34/74). The error rates for the 6 MRAs differed significantly ($p < .05$); the lowest was 8.0%, the highest 25.6%.

We also evaluated the MRA screening process in light of the AEAF completed by the senior physicians. The MRA screening process failed to identify positive screening criteria in 19 out of a total of 241 reviews that were positive for causation (7.9%). MRAs failed to identify positive screening criteria in only 9 out of 144 reviews positive for negligence (6.3%). Using the estimate of 7.9% for sensitivity of the screening process, an estimate of 25% for the prevalence of charts referred in the screening process from our review of 30,000 medical records in New York State, and the estimate of 5% for the incidence of adverse events from the Medical Insurance Feasibility Study, we calculated the negative predictive value of the screening process to be 99.5% and positive predictive value to be 21%, (Table 2).

B Physician-Reviewer Reliability Study

Ten physician-reviewers completed 875 of a possible 900 reviews (97.2%). The reliability (R_m) for the average of the causation scores provided by two physician-reviewers ($m= 2$) using the 10-point scale was .78, suggesting excellent reliability. The binary causation scale also produced good to excellent reliability when based on the average of two physician-reviews (Table 3). Judgments on negligence were less reliable than those of causation but were well within the "good" range (Table 3).

A jackknife 95% confidence interval for intra-class correlation coefficient of causation on the 10-point scale for a single reviewer was (.56, .72), and the corresponding interval

for R_m (m=2) was (.72, .84). For the average of two raters, therefore, the confidence interval for causation lies solidly in the good-to-excellent range. This analysis also revealed the impact of each physician-reviewer on the overall reliability figure on the ten point scale for causation (Table 3). For instance, omitting physician-reviewer #3 increased R to .65, indicating his reviews tended to be less reliable when compared to other reviewers.

As noted, we adjusted our reliability estimates in light of the non-random sampling strategy we employed. Using the binary scale, the odds ratio for repeated reviews for any pair of reviewers was estimated to be 15.0 for causation and 7.2 for negligence. These large odds ratios reflect the very high association between ratings by pairs of reviewers. Records were reviewed by physicians only if they satisfied one of the 16 criteria. Among this group we anticipated an **adverse event** rate of approximately 0.2. Negligence was determined only for those records with any evidence of causation; we anticipated the prevalence of negligence in this group to be 0.1. Using these assumptions and the estimated odds ratio, we computed the adjusted reliability (again based on two reviews) to be .68 for causation and .44 for negligence.

C. Senior Physician Correlation Study

A comparison of senior physician consensus causation scores and the mean causation score for 4 physician-reviewers, both based on the binary scale, revealed moderate to good correlation of the two processes (kappa = .57). The same was true with regard to negligence (kappa = .51).

To assess the impact of difficult cases on the reliability of physician-reviewer judgments, we calculated separately the

reliability for difficult and other cases. As we expected, the causation judgments were far less reliable for difficult cases (R_m= .48, m=1) than for other cases (R_m=.65, m=1). There was even less reliability on the judgments concerning negligence on the difficult cases (R_m=.11, m=1) (other cases' reliability on negligence, R_m=.47, m=1).

These results demonstrate that physician judgments were reliable and valid and that the MRA screening process was quite valid as well. Much of this technical appendix is reprinted from Medical Care, December 1989 with permission.

Table 1
Comparison of MRA Reviews with Senior Physicians' Reviews on Screening Criteria

		Senior Physicians		
MRA	Criteria Not Present	140	74	214
	Criteria Present	55	403	458
		195	477	672

Sensitivity = 84.5%
Specificity = 71.8%

Table 2

Estimate of Screening Process Performance in Light of Physician Judgments Regarding Adverse Events

		Physician Judgment About Adverse Event AE Absent	AE Present
MRA Screening Process	Criteria Absent	.7461	.0039
	Criteria Present	.204	.046

$$\text{Postive Predictive Value} = \frac{.046}{.204 + .046} = .23 = .184$$

$$\text{Negative Predictive Value} = \frac{.7461}{.7461 + .0039} = .995$$

Table 3

Reliability of Judgments Concerning Causation and Negligence Using Different Scales

Dependent variable	R	Rm (m=2)
Cause:		
Cause on 10-point scale	.63	.78
Cause on binary scale	.57	.72
Negligence:		
Negligence on 10-point scale	.45	.62
Negligence on binary scale	.34	.51

MEDICAL PRACTICE STUDY:
ADVERSE EVENT ANALYSIS FORM

Directions:

Complete this form only if the record review by medical record administrators or other staff has found that a screening criterion is fulfilled.
DO NOT WRITE MEDICAL RECORD NUMBERS ON THIS FORM.

Place your answers in the space or box beside the appropriate section.
PLEASE PRINT OR WRITE RESPONSES OR NOTES LEGIBLY.

"AE" means adverse event.

CASE NO. | | | | | | | |

(Copy from the top of the 3-page Hospital Record Screen)

IF FOUND, PLEASE RETURN TO DIRECTOR OF MEDICAL RECORDS

1. Reviewer's ID number: ☐☐☐

2. Date this review: ☐☐ ☐☐ 88
mm dd yy

3. DO YOU CONFIRM THE MEDICAL RECORD ADMINISTRATOR'S POSSIBLE AE FINDING?

1 ☐ Yes
0 ☐ No

IF YOU ANSWERED "YES" TO Q. 3 GO TO Q. 4, *DETERMINATION OF ADVERSE EVENT*

3.1. IF YOU ANSWERED NO TO Q. 3, INDICATE IN BOX AT RIGHT THE REASON FOR YOUR ANSWER.

1. criterion error (criterion not fulfilled, no AE)
2. criterion inappropriate (criterion fulfilled, no AE);
   Explain: _____
   _____
3. inconsequential finding (no prolongation of hospitalization, no significant disability at discharge)
4. inadequate documentation (causation indeterminate due to deficiencies in record)
5. other (specify) _____
   _____

SCORE

IF NO POSSIBLE AE FOUND, STOP: END OF ADVERSE EVENT FORM, THANK YOU.

THIS SPACE FOR MPS STAFF USE ONLY

4. DETERMINATION OF ADVERSE EVENT

In your best judgment, is there evidence that medical management caused the patient's injury? In answering this question, consider when relevant the following questions and CHECK THE APPROPRIATE BOXES IN COLUMN AT RIGHT:

4.1. Is there a note in the medical record which indicates or suggests that medical management caused the injury?	1 ☐ **No (suggests no AE)** 2 ☐ **Yes (suggests AE)**
4.2. Is there a note in the medical record which predicts the possibility of an injury from the patient's disease?	1 ☐ **Yes (suggests no AE)** 2 ☐ **No (suggest AE)**
4.3. Does the timing of events suggest that the injury was related to the treatment?	1 ☐ **Unlikely** 2 ☐ **Possibly** 3 ☐ **Likely**
4.4. Are there other reasonable explanations for the cause of the injury?	1 ☐ **Many competing explanations** 2 ☐ **Some competing explanations** 3 ☐ **Few competing explanations**
4.5. Was there an opportunity prior to the occurrence of the injury for intervention which might have prevented it?	1 ☐ **No (suggests no AE)** 2 ☐ **Possibly** 3 ☐ **Yes (suggests AE)**
4.6. Is there recognition that the intervention in question causes this kind of an injury?	1 ☐ **No** 2 ☐ **Recognized by specialists** 3 ☐ **Widely recognized**
4.7. Did the adverse event respond to new management to neutralize or modify effects of former management?	1 ☐ **No such response (suggests no AE)** 2 ☐ **Suggestive response** 3 ☐ **Convincing response (suggests AE)**

CONSIDER EACH OF THE ABOVE QUESTIONS PRIOR TO CONTINUING WITH Q. 4

4.8. After due consideration of the clinical details of the patient's management, AND YOUR RESPONSES TO Q. 4.1-4.7, what level of confidence do you have that the MEDICAL MANAGEMENT CAUSED THE INJURY? (Indicate one score in box at right)

Confidence Score

1. Little or no evidence for management causation
2. Slight to modest evidence for management causation
3. Management causation not quite likely; less than 50-50 but close call
4. Management causation more likely than not; more than 50-50 but close call
5. Strong evidence for management causation
6. Virtually certain evidence for management causation

SCORE ☐

AE DESCRIPTION:

4.91. Please describe briefly the adverse event (not how it occurred):

_____

_____

_____

_____

_____

4.92. Please describe briefly the relation of AE to medical management and the disease process:

_____

_____

_____

_____

_____

IF YOUR CONFIDENCE SCORE FOR CAUSATION IN 4.8 IS 2 OR GREATER, THEN PROCEED TO QUESTION 5.1, *SPECIAL SITUATIONS*. IF YOUR CONFIDENCE SCORE IS 1, THEN PROCEED TO QUESTION 11.

5.1. Special situations (Check all that apply):

Was the AE:

1 ☐	a drug reaction or drug side effect?
2 ☐	a fall?
3 ☐	a wound infection?
4 ☐	the result of a failure/delay in diagnosis?
5 ☐	the result of an omission of appropriate therapy?
6 ☐	the result of experimental medical treatment approved by a human subjects committee?
7 ☐	the result of a defective device or product?
8 ☐	involving a DNR (do not resuscitate) patient?

IF YOU CHECKED 1, DRUG REACTION OR SIDE EFFECT, THEN ANSWER Q. 5.21-5.24

5.21. Was the drug (Indicate one drug in box at right):

1. antibiotic
2. antineoplastic agent
3. anti-seizure medication
4. diabetes drug
5. cardiovascular agent
6. asthmatic agent
7. sedative, hypnotic
8. peptic ulcer medication
9. antihypertensive
10. antidepressant
11. antipsychotic
12. analgesics
13. other (specify)_____

DRUG ☐

5.22. What was the drug? _____

5.23. What was the side effect? _____

5.24. Would a physician using reasonable medical judgment prescribe the drug even with knowledge beforehand that this side effect would occur?

1 ☐ Yes
0 ☐ No

PATIENT CONTRIBUTION TO AE:

5.31. Is there evidence that patient conduct contributed in important measure to this AE?

1 ☐ Yes
0 ☐ No

IF YOU HAVE ANSWERED "NO" TO Q. 5.31 THEN SKIP TO Q. 6.1, *DISABILITY*

5.32. Is there evidence that the patient was not competent to make rational decisions with regard to medical therapy or management?

1 ☐ Yes
0 ☐ No

IF YOU HAVE ANSWERED "YES" TO Q. 5.32 THEN SKIP TO Q. 6.1, *DISABILITY*

5.33. If you have answered YES to Q. 5.31 and NO to Q. 5.32 THEN rate on a 6 point scale your confidence in your judgment that the patient contributed to the AE. (Indicate one score in box at right)

Confidence Score

1. Little or no confidence
2. Slight to modest confidence
3. Patient contribution not quite likely; less than 50-50 but close call
4. Patient contribution more likely than not; more than 50-50 but close call
5. Strong confidence
6. Virtually certain

SCORE ☐

6. DISABILITY

6.1. Was the patient's length of hospitalization extended because of this adverse event?

1 ☐ Yes
0 ☐ No

6.15. If YES, estimate the number of days of increased hospital stay because of the adverse event.

DAYS ☐☐☐

6.2. Check the organ system(s) or bodily function(s) involved as part of the AE. (Check all that apply)

- 6.2a ☐ central nervous/cognitive
- 6.2b ☐ central nervous/sensory
- 6.2c ☐ central nervous/motor
- 6.2d ☐ central nervous/psychiatric
- 6.2e ☐ peripheral nervous/sensory
- 6.2f ☐ peripheral nervous/motor
- 6.2g ☐ cranial nervous/sensory
- 6.2h ☐ cranial nervous/motor
- 6.2i ☐ vision
- 6.2j ☐ smell
- 6.2k ☐ taste
- 6.2l ☐ hearing
- 6.2m ☐ speech (larynx, tongue, vocal cords)
- 6.2n ☐ blood forming/hematologic
- 6.2o ☐ immune system
- 6.2p ☐ cardiovascular/circulation
- 6.2q ☐ endocrine (including breasts)
- 6.2r ☐ gastrointestinal
- 6.2s ☐ reproductive or genital
- 6.2t ☐ urinary/renal
- 6.2u ☐ respiratory
- 6.2v ☐ skin
- 6.2w ☐ musculoskeletal
- 6.2x ☐ other (specify) _____

6.3. Degree of patient's disability caused by this AE over and above patient's disability from underlying disease. (Indicate one score in box at right)

1. Minimal impairment (functional, cosmetic) followed by almost complete recovery **within 1 month** of sustaining AE.

2. Moderately incapacitating impairment (functional, cosmetic) followed by almost complete recovery in more than 1 month but **within 3 months** of sustaining AE.

3. Same as 2 with almost complete recovery in more than 3 months but **within 6 months** of sustaining AE.

SCORE

4. Moderately incapacitating impairment (functional, cosmetic) followed by almost complete recovery in **more than 6 months** or incomplete recovery not interfering significantly with employment or leisure activity.

5. Disability causing **1% to 50% permanent** decrease in employment or leisure activity.

6. Disability causing **51% to 100% permanent** decrease in employment or leisure activity.

7. Requires personal and/or nursing **support permanently**.

8. **Death**

9. Cannot reasonably judge disability from medical record.

6.41. WHEN did medical management and the consequent AE occur in relation to the index hospitalization. (Indicate one score in box at right)

1. Medical management **during** index hospitalization; AE discovered **during** index hospitalization.

2. Medical management **during** index hospitalization; AE discovered **after** discharge and **during outpatient treatment** (e.g. ER visit which might or might not lead to hospitalization)

SCORE

3. Medical management **during** index hospitalization; AE discovered **after** discharge and **during subsequent hospitalization**.

4. Medical management in **outpatient treatment prior** to index hospitalization (but after Jan. 1, 1981); AE indication for index hospitalization or first discovered **during** index hospitalization.

5. Medical management during **hospitalization** at any institution **prior** to index hospitalization (but after Jan. 1, 1981); AE indication for index hospitalization or first discovered **during** index hospitalization.

6.42. If causal medical management occurred prior to the index hospitalization, give date of management: (e.g., 01/83 for January 1983)

mm yy

6.43. If AE discovered after the index hospitalization, give date of AE:

mm yy

6.5. WHERE did the medical management event causing the AE occur? (Indicate one site in box at right)

SITE

Outside Hospital

1. Physician's office
2. Ambulatory care unit (including day surgery) outside hospital
3. Home
4. Home, labor and delivery
5. Nursing home
6. Other site outside of hospital (specify) _____

In Hospital

7. Patient's hospital room
8. OR
9. ICU
10. ER
11. Ambulatory care unit (including day surgery) inside hospital
12. Recovery room
13. Labor and delivery
14. Nursery
15. Radiology
16. Cardiac catheterization
17. Therapy/rehabilitation
18. Pathology
19. Laboratory (clinical)
20. Blood bank
21. Pharmacy
22. Hospital bathroom
23. Service areas (stairs, halls, elevator)
24. Procedure room
25. Other site in hospital (specify) _____

6.6. SPECIALTY or category responsible for AE (Indicate one specialty from the following list in box at right):

SPECIALTY

Surgery

1. anesthesiology
2. cardiac surgery
3. colon/rectal surgery
4. general surgery
5. gynecology
6. neurosurgery
7. obstetrics
8. orthopedic surgery
9. pediatric surgery
10. plastic surgery
11. thoracic surgery
12. vascular surgery
13. urological surgery

Medicine

14. cardiology
15. dermatology
16. emergency medicine
17. endocrinology
18. family practice
19. gastroenterology
20. hematology
21. immunology and allergy
22. infectious diseases
23. intensivist (ICU)
24. internal medicine (not otherwise classified)
25. medical oncology
26. neonatology
27. nephrology
28. neurology
29. ophthalmology
30. otorhinolaryngology (ENT)
31. pathology
32. pediatrics
33. physical medicine
34. psychiatry
35. pulmonary disease
36. radiation therapy
37. radiology (diagnostic)
38. rheumatology
39. other (specify)_____

Non MD Specialty and Other Categories

40. dentistry/oral surgery
41. dietary
42. hospital physical plant
43. midwifery
44. nursing
45. osteopathy
46. pharmacy
47. physical or occupational therapy
48. podiatry
49. transportation support services
50. other (specify)_____

7. IS THERE EVIDENCE FOR NEGLIGENCE?

1 ☐ Yes
0 ☐ No

7.1. Was this AE possibly due to a reasonably avoidable error, or carelessness by either an individual or medical care system, or both?

IF YOUR ANSWER IS NO, GO TO Q. 11. OTHERWISE CONTINUE WITH Q. 7.2.

7.2. CLASSIFY the error. In a few cases more than one error might be present. If that is the case, RANK IN ORDER OF IMPORTANCE (1=most important).

Rank	
7.21 ☐	Error in DIAGNOSIS including failure to order or perform appropriate diagnostic procedures or tests.
7.22 ☐	Error in PREVENTION of adverse event due to failure to employ accepted standards of management.
7.23 ☐	Error in PERFORMANCE OF A PROCEDURE OR OPERATION.
7.24 ☐	Error in DRUG TREATMENT.
7.25 ☐	SYSTEM ERROR primarily caused by the medical care system.
7.26 ☐	OTHER ERROR not related to above (specify)_____

7.3. If the AE resulted from a DIAGNOSTIC error, was it due to (Check as many as apply):

7.31 ☐	Failure to employ indicated tests?
7.32 ☐	Failure to act upon results of tests or findings?
7.33 ☐	Use of inappropriate or outmoded diagnostic tests?
7.34 ☐	Avoidable delay in diagnosis?
7.35 ☐	Physician or other professional practicing outside area of expertise?
7.36 ☐	Other diagnostic error (describe): _____
7.37 ☐	Reason for error not apparent?

7.4. If the AE resulted from an error in PREVENTION, was it due to (Check as many as apply):

7.41 ☐	Failure to take precaution to prevent accidental injury?
7.42 ☐	Failure to employ indicated tests?
7.43 ☐	Failure to act upon results of tests or findings?
7.44 ☐	Use of inappropriate or outmoded diagnostic tests?
7.45 ☐	Avoidable delay in treatment?
7.46 ☐	Physician or other professional practicing outside area of expertise?
7.47 ☐	Other prevention error (describe):_____

7.5. If the AE resulted from an error in the PERFORMANCE OF A PROCEDURE OR OPERATION, was it due to (Check as many as apply):

7.51 ☐	Inadequate preparation of patient before?
7.52 ☐	Technical error?
7.53 ☐	Inadequate monitoring of patient afterwards?
7.54 ☐	Use of inappropriate or outmoded form of therapy?
7.55 ☐	Avoidable delay in treatment?
7.56 ☐	Physician or other professional practicing outside area of expertise?
7.57 ☐	Other performance error (describe): _____

7.6. If the AE resulted from an error in DRUG TREATMENT, was it due to (Check as many as apply):

7.61 ☐	Error in dose or method of use?
7.62 ☐	Failure to recognize possible antagonistic or complementary drug-drug interactions?
7.63 ☐	Inadequate follow-up of therapy?
7.64 ☐	Use of inappropriate drug?
7.65 ☐	Avoidable delay in treatment?
7.66 ☐	Physician or other professional practicing outside area of expertise?
7.67 ☐	Other drug treatment error (describe): _____

_____

7.7. If the AE resulted from a SYSTEM ERROR, was it due to (Check as many as apply):

7.71 ☐	Defective equipment or supplies?
7.72 ☐	Equipment or supplies not available?
7.73 ☐	Inadequate monitoring system?
7.74 ☐	Inadequate reporting or communications?
7.75 ☐	Inadequate training or supervision of MD or other personnel?
7.76 ☐	Delay in provision or scheduling of service (Xray, lab tests, follow-up visit, etc.)?
7.77 ☐	Inadequate staffing?
7.78 ☐	Inadequate functioning of hospital service (e.g., pharmacy, blood bank, housekeeping)?
7.79 ☐	Other system error (describe): _____

_____

8. If any of the following possible MITIGATING or AGGRAVATING CIRCUMSTANCES were present, CHECK ONE STATUS IN BOXES AT RIGHT FOR EACH CIRCUMSTANCE.

8.1 Degree of deviation of treatment from accepted norms	1 ☐	little
	2 ☐	moderate
	3 ☐	severe
8.2 Degree of additional morbidity created for typical patient with similar age, diagnosis, and severity of illness, who is victim of this kind of negligence (i.e., not the specific degree of morbidity in the patient under consideration)	1 ☐	little
	2 ☐	moderate
	3 ☐	severe
8.3 Number of all patients who are at risk for this kind of negligence	1 ☐	small
	2 ☐	moderate
	3 ☐	large
8.4 Degree of emergency in management of case prior to occurrence of negligence	1 ☐	critical & very urgent
	2 ☐	moderate
	3 ☐	not urgent
8.5 Complexity of case in which negligence occurred	1 ☐	very complex
	2 ☐	moderately complex
	3 ☐	uncomplicated
8.6 Co-morbidity of case in which negligence occurred	1 ☐	very ill patient
	2 ☐	moderately ill patient
	3 ☐	no co-morbidity
8.7 Lack of consensus about correct therapy or diagnosis (even among experts)	1 ☐	very little consensus
	2 ☐	some consensus
	3 ☐	great deal of consensus

8.8. After having considered the factors in 8.1-8.7 you might have reassessed whether negligence occurred. If you feel there is NO negligence, CHECK THE SPACE ON THE RIGHT AND GO TO Q. 11.

NO NEGLIGENCE ☐

8.9. If you find that negligence did occur, what is the severity of this type of negligence? (Indicate one score in box at lower right)

Severity Score
1. Slight
2. Moderate
3. Grave

SCORE ☐

9. If you have given a Severity Score of 1, 2, or 3, then rate on a 6 point scale your confidence in the evidence for negligence. (Indicate one score in box at right)

Confidence Score

1. Little or no evidence for negligence
2. Slight to modest evidence for negligence
3. Negligence not quite likely; less than 50-50 but close call
4. Negligence more likely than not; more than 50-50 but close call
5. Strong evidence for negligence
6. Virtually certain evidence for negligence

SCORE ☐

10. PLEASE DESCRIBE BRIEFLY THE NEGLIGENCE:

_____

_____

_____

_____

_____

ADDITIONAL AEs:

11. If you have identified an AE, is there an additional AE that contributed to the patient's disability, is attributable to provider negligence, and is not described in your responses to questions 1-10?

1 ☐ Yes
0 ☐ No

11.1 If YES, describe briefly the medical management, the AE, and the negligence:

_____

_____

_____

_____

_____

12. Reviewer's judgments limited or hampered by lack of subspecialty knowledge. (Check "Yes" if you think a specialist's review is necessary)

1 ☐ Yes
0 ☐ No

12.1 If YES, which specialty? (List as many as necessary)

_____ _____ _____

12.2 Describe the judgment which is limited or hampered by lack of subspecialty knowledge and the clinical question you would pose to a specialist.

_____
_____
_____
_____
_____
_____
_____
_____
_____
_____

12.3 Describe the resolution of the question(s) posed following consultation with a specialist.

_____
_____
_____
_____
_____
_____
_____
_____
_____
_____

12.4 Specialist's ID number: | | | |

12.5 Specialist's ID number: | | | |

END OF AE ANALYSIS, THANK YOU

DISCOVERABILITY OF ADVERSE EVENTS THROUGH A RECORD REVIEW

In order to determine the utility of the medical record review process to uncover **adverse events** (AE) suffered by hospitalized patients as well as substandard care, we reviewed records from two metropolitan teaching hospitals, making comparisons to risk management and litigation files. We used methods described in Technical Appendix 5.IV.1 to identify **adverse events** and negligence, or substandard care, in the medical records. We also examined hospital quality assurance files to assess the efficacy of current programs to identify and address the care that gave rise to the adverse events.

I. Methods

A. Sample Selection

We selected records from the malpractice claims files of the two hospitals to develop a study sample with a sufficient number of **adverse events** and to compare physician reviews of records with reviews of litigation records. A claim file exists because:

(1) the patient filed a malpractice suit (suit);
(2) the patient did not file a suit but sought monetary damages from the hospital or physicians (claim); or
(3) the hospital's risk management office considered a medical event to be serious enough to warrant an investigation (observation).

Therefore, the records reviewed had a much higher incidence of **adverse events** than the hospital population.

The hospitalization that gave rise to a claim was designated the **index hospitalization** and the corresponding record was studied. Control records from the two hospitals in 1984-

1987 were randomly selected from patients with similar lengths of stay. A total of 93 cases and 100 controls in hospital A, and 62 cases and 72 controls in hospital B had complete reviews. Cases and controls were inter-mixed and reviewers were blind to their status.

B. Identification of Adverse Events

Our method for identifying **adverse events** was an enhancement and standardization of the two-stage process employed by Mills et al. in the Medical Insurance Feasibility Study and is described in Technical Appendix 5.IV.1.

C. Insurance Records Review

The two study hospitals and most of their physicians obtain liability insurance through a single off-shore captive insurer via a central risk management office. A risk management coordinator at each hospital forwards information on suits, claims, and observations to the central office, which maintains complete records of all cases. One of us, trained in both law and internal medicine, reviewed the medical chart and the claims file at the central office and classified each case according to level of activity (observation, claim or suit), disposition and payment.

In consultation with another of the surgeon-investigators, the lawyer-internist used an abbreviated AEAF to grade the records for causation of **adverse event** and negligence based on the information available from the litigation records. These scores were eventually collapsed into a binary score. Those cases that led to **adverse events** that were impossible to discover by record review at the index hospital because the **adverse event** was first discovered at a subsequent hospitalization in another

institution were also noted as were other categories of cases. Several case studies characterizing the false negatives were developed.

D. Statistical Analysis

We used a kappa statistic to compare scores from the physician review and the review of risk management records. In addition, the subsets of cases involving suits or claims were examined separately. Sensitivity and specificity of the physicians' review process were calculated using the litigation score as the gold standard.

E. Confidentiality

MRAs, physician-reviewers and senior physicians were instructed repeatedly in the importance of confidentiality. All data collection devices were referenced by code number specific only to this study. The risk management abstracts, as well as all the identification codes, were locked in a safe at a central office. After completion of the data collection and analysis, medical record identifiers were destroyed.

II. Results

The 155 risk management records provided us with information on the sensitivity and specificity of physician-based medical record reviews aimed at uncovering **adverse events** and negligence. With regard to **adverse events**, in 23/116 records in the medical record review missed an adverse event that was identified in the review of litigation records. Thus the sensitivity is 80% (Table 1). The sensitivity for negligence is somewhat lower (51/67,

76.1%), indicating that evidence of negligence is less readily available from a review of medical records than is evidence of **adverse events** (Table 2). The overall kappa was .57 for causation and .62 for negligence, both within the range of good agreement.

Further insight into the limitations of a record review can be gleaned by classifying the reasons for a missing adverse event. Five **adverse events** were missed because the adverse event was discoverable only when the patient was admitted to a different hospital. In a random sample of hospitalizations at the state or federal level, these **adverse events** would be discovered. Thus the sensitivity of a review of a random sample of hospitalizations is 83.8% (negligence 80.1%). Other reasons for failing to identify **adverse events** are listed in Table 3. Notably, the overwhelming majority were due to physician-reviewer error rather than deficiencies in the record. Of those that were not discoverable, consider the following illustrative case studies:

Case One. Two days following surgery a patient receiving ventilatory assistance experienced a fatal cardiac arrest. Subsequent investigation by hospital staff revealed that the respirator had malfunctioned. This malfunction was not recorded in the medical chart and thus could not be discovered in the record review process. Information not in record, not discoverable.

Case Two. An elderly patient who had recently suffered a cerebrovascular accident fell out of bed and sustained a fracture. The medical record did not mention that a hospital risk management investigation revealed that the protective side rails were broken. Therefore, the reviewers were

unable to identify the negligence. Information not in record, not discoverable.

Case Three. A young woman was admitted to the hospital for overnight observation because of abdominal pain. She was discharged the following day with no evidence of positive findings in the record except for a significant elevation in the white blood cell count and a mild fever. She was admitted to another hospital later that day where she was treated for a ruptured appendix. Information not in record, but discoverable on survey of random sample of hospitalizations.

Analysis of those claims for which the litigation process is completed (closed claims) casts further light on the integrity of the record review. In terms of payment, looking at the severity of the false-negative cases for causation, we found that only 1/17 had resulted in payment, and this in the $5,000-$25,000 range. In contrast, if we define the review of risk management records as the gold standard, then among the true positives for causation, 13/77 resulted in payment (3 for less than $5,000, 6 for $5,000-$25,000, 2 for $25,000-$100,000, 1 for $100,000-500,000, and 1 for more than $500,000). Thus those claims identified in the record room as containing **adverse events** are far more likely to result in payment. Conversely, those claims not classified by a record review as **adverse events** are much less likely to result in payment.

This study also provides some insights into the prevalence of **adverse events** in hospitalized patients. Among the 172 control records, 20 hospitalizations that were not the subject of risk management review or litigation nonetheless contained **adverse events** on physician review of the medical record. Six of the **adverse events** were classified by physician-reviewers as

caused by negligence. These **adverse events** gave rise to disabilities ranging from minor to death (Table 4). One, concerning a negligent failure to diagnose colon cancer, could not be rated according to disability.

These results indicate that the overwhelming majority of **adverse events** are discoverable through a record review.

Table 1: Identification of Adverse Event: Senior Physician Review of Risk Management Records

		Litigation Record Review	
		AE Absent	AE Present
Senior Physician Review	AE Absent	33	23
	AE Present	6	93

Sensitivity - 93/116 = 80%
Kappa = .57

Table 2: Identification of Negligence: Senior Physician Review of Risk Management Records

		Litigation Record Review: Negligence Absent	Litigation Record Review: Negligence Present
Senior Physician Review	Negligence Absent	75	16
	Negligence Present	13	51

Sensitivity - 51/67 = 76%
Kappa = .62

Table 3: Reasons For Failing to Identify Adverse Events on Record Review

Reason	Frequency	
	Adverse Events n = 23	Negligence n = 16
Reviewer Error	14 (61%)	6 (37%)
Adverse Event Discoverable At Different Hospital	5 (22%)	3 (19%)
Adverse Event Discoverable at Out-Patient Visit	1 (4%)	0 (0%)
Information not in Record	3 (13%)	7 (44%)

Table 4

CONTROL RECORD
ADVERSE EVENT DISABILITY RATINGS (20/172)

Status	Frequency	Disability Rating
Adverse Event	4	Minor, Temporary
Adverse Event	9	Major, Temporary
Adverse Event	1	Death
Negligent Adverse Event	1	Minor, Temporary
Negligent Adverse Event	3	Major, Temporary
Negligent Adverse Event	1	Major, Permanent
Negligent Adverse Event	1	Not Rated

DATA MANAGEMENT, VERIFICATION, AND IMPUTATION

I. Data Management in Field

This section discusses the data collection instruments used in the field during the medical record review. These instruments include: the Medical Record List, the Hospital Record Screen, the Leader Master Worksheet, and the Discrepant Review Form.

A. Management by Medical Record Administrator (MRA) Leader

Each hopsital team has a designated leader (MRA Leader), who used a Master Worksheet for monitoring the progress of the review, identifying all reviewers, tracking records, and documenting unavailability of a medical record during the time of review. The Worksheet, with preprinted medical record identifiers and a confidential Study-supplied case identifier, and blank spaces for the Leader to complete, enabled the record review team to locate and abstract the sampled medical records.

The MRA Leader used the Worksheet to document the MRA's personal identification code and the date of review. It was noted if the case was referred for physician review, and if so, the identification codes for the physician-reviewers for the case were recorded. Additionally, whether the physician reviews agreed or disagreed were noted.

If the record room could not locate a record, the Leader noted the dates of the first, second and third requests, the final date the record was declared not available, and the reason the record was unavailable. The Worksheet then became the primary document on the entire first-stage review process.

1. Hospital Record Screen. The Hospital Record Screen, a 3-page form used by the MRA for each medical record, was preprinted with available demographic, billing, and hospital admission

information for identification of each sampled hospitalization. The third page of the screen contained the 17 criteria with ample space for notations under each. The MPS provided the MRA Leader for each hospital with these three-page forms on one continuous strip of fan-fold paper. Each set of 3 pages was to be separated by the MRA Leader and given to the MRA for the first-stage review.

The Leader checked that each Hospital Record Screen included the name of the patient, whether or not any criterion was met. If a criterion was met, then the Leader obtained all the necessary identification information for locating patients for possible interview. (see Chapter 8)

2. Discrepant Review Form. In the event of a discrepancy between physician reviews for a case, the Leader noted for the Physician-Supervisor (PS) on a Discrepant Review Form the case identifier and the items in disagreement.

B. Record Return To Central Office

To avoid losing data that could not be reproduced, all forms and logs were returned in person to the Study's central offices by a courier, who was under agreement with the Study to preserve the confidentiality of all information.

II. Data Management At Study Office

Upon return to the central office, the cases were coded as complete on a copy of the Leader's Log. The forms were then reviewed for any potential problems during for the keypunching process. With all problems solved, the forms were ordered and sent out to be keypunched. After keypunching, the information for each case (Hospital Record Screen, the Adverse Event Analysis Form, and the Discrepant Review Form) was filed and stored.

A. Initial Data Verification

Upon return from the field to the Medical Practice Study, all data collection forms were sorted for initial review against the Leader Worksheet as to the status of the case (i.e., not referred for review, referred, missing medical record, etc.). All forms (Hospital Record Screen, Adverse Event Analysis Form, Discrepant Review Form, and the Leader Log) were checked for legibility and for valid responses to each question. All changes were noted in red pencil in order to preserve a record of clerical changes prior to keypunching.

1. **Hospital Record Screen.** All variables were reviewed for consistency and valid repsonses. For example, unnamed newborns were often noted in different fashions by the MRAs, and these were coded for consistency as "baby boy" or "baby girl." Valuable information noted under any of the criteria on the Hospital Record Screen was recorded by the Study, when appropriate, under "other patient information" (e.g., "death of the patient in 1987" noted under one criterion was also recorded by the Study as "death post-discharge" under "other patient information"). Missing information, such as area code and zip code, was obtained from directories.

2. **Adverse Event Analysis Forms.** For each physician review, we verified the case number for validity and checked it against the Leader Log to confirm that the review was done by the physician-reviewer whose identity code appeared on the form. Once again, all responses were pre-checked for legibility for keypunching.

Incorrect coding of answers occurred for a number of items on this form. One variable, concerning error classification (Q.7.2), directed physicians to rank their responses in order of importance. When this direction was not followed and the responses were simply checked off without ranking, the responses were then coded as "9" to indicate unranked responses. We also found responses to the specific questions in Q7.3-7.7 were answered but the corresponding general question (Q7.2) was not answered. For these cases, we completed Q.7.2 with the value of "9" (unranked) or "1" if all of Q.7.2 had been blank, so that future computer retrieval could track the responses.

Some physicians had also left self-coding responses blank but had responded to related questions (e.g., Q.5.21-5.24 concerning drug reactions and side effects were answered but Q.5.11 was left blank). In these situations, we coded an appropriate response. In a few cases, classification of a drug was left blank by the physician-reviewer even though the drug was named and the side effect listed. These cases were referred to a senior physician for drug classification (Q5.21).

Errors in the timing and the relationship of medical management to the adverse evemt (AE) were reviewed in the field by senior physicians when the two physician-reviewers disagreed. Periodically, however, both reviewers coded the answers to these questions similarly but inaccurately. By reviewing the Hospital Record Screen, the AE description, and the relationship of medical management to the AE, it was often possible to determine the correct responses to Q.6.41-6.43.

A number of variables required the reviewer to indicate only one answer in the box provided. If more than one response was noted, we classified the response as unknown. We found this

error most often in Q.6.6 (specialty responsible for the AE) where a reviewer incorrectly indicated multiple responses.

3. **Discrepant Review Forms.** The Discrepant Review Form was also checked for correct case number, physician-supervisor and reviewer ID numbers, and action taken. In several instances in which an ID number was incorrect (no review with the indicated ID number for that case), we corrected the review by consulting the Leader Worksheet (see above).

Following this initial review and cataloguing, the data was keypunched onto magnetic tape or disk and verified.

B. Secondary Data Verification

Prior to data analysis, we further verified all keypunched data by computer. First, each review was assessed for errors within the review. All case numbers were checked and determined to be valid. The number of physician reviews for each case was checked. Cases containing key questions that were left blank were checked. Once again, Q.6.41-6.43 (timing and relationship of medical management and the AE) were checked for errors. And finally, errors in ranking (Q.7.2) were noted in a listing.

A second program checked for discrepancies between reviews for each case. This program converted appropriate missing values to "0" or another appropriate indicator. For factual responses (e.g., Q. 6.41-6.43 the timing of medical management), this verification ensured that physicians were reporting the case in the same way. Because these are factual items, when differences were noted it was possible to determine the correct responses by

reviewing the Hospital Record Screen along with the Adverse Event Analysis Forms for the particular case. Date fields were also checked for range. For example, answers to Q.6.42 had to fall within the time frame of 1/1/81 to 12/31/84, while dates for Q.6.43 had to be greater than 12/31/83.

C. Problem Cases

1. **Number of Reviews for a Case.** As part of the secondary validation of keypunched cases, we ascertained the number of reviews completed for each case. In a few cases, too many reviews were performed. In this situation, we relied on the date of the reviews and used the earlier two reviews, under the assumption that after the first two reviews, all later reviews were unnecessary and therefore invalid.

In a few cases, two reviews were performed for the same case by the same physician-reviewer. In these situations, the review with the earlier date was kept and the other was discarded.

In 33 cases, only one physician review was obtained instead of two, most likely due to the administrative complexity of scheduling physicians, record room personnel, and hospital leaders. These cases were judged on the results of the single review and flagged as having only one review.

2. **Discrepant Reviews.** Via the secondary data validation step, 51 cases were found to be discrepant and in need of a review by a senior physician per the Study protocol. Evidently, these cases were not identified as such in the field. Ultimately, two types of discrepant reviews needed resolution.

The first type concerned the presence or absence of an AE and occurred 31 times. In this situation, one physician found an AE and confirmed the MRA's findings (answering Q.3 "yes"), while the other reviewer found no AE and answered Q.3 "no." Without a senior physician review to side with the No AE reviewer and reject the AE review, or vice versa, we developed an alternative resolution process. For all discrepant reviews of this nature that were solved during the record review, the physician supervisors found AEs in 27% of the cases and no AEs in 73%. Thus, we chose to resolve discrepancies by selecting (with probability =.27) the physician review with an AE. This method resulted in our picking 10 AE reviews (32.3%) and 21 non-AE reviews (67.7%). The superceded reviews were no longer used in any analyses.

The second type of discrepancy consisted of two different AEs for one case or different timing of medical management. In 5 of 20 such cases, a senior physician determined retrospectively that the reviewers were describing essentially the same AE. Both reviews were kept, after correcting any timing errors and were then removed from the problem-case classification.

Twelve of the 20 cases involved two clearly different AEs being described. Determining which reviewer to discard depended on his history of being overruled by a senior physician in the field. Thus, if the probability that the first reviewer was accepted by a senior physician was pr (acc_1) and the probability that the second reviewer was accepted was pr(acc_2),

then PR_1 was chosen randomly with $= \frac{pr(acc_1)}{(pr(acc_1)+pr(acc_2))}$.

Three cases differed in timing of medical management and discovery of the AE (Q.6.41). With the methods described above, we chose the timing of one review randomly and used that timing

for both reviews. The remaining responses of both reviews were not altered.

III. Imputation for Item Non-response

Through the secondary data verification process, 40 cases were found to have item non-response (at least one question that should have been answered was not on one or both reviews). We want to emphasize that 39 of these 40 cases had two reviews and at least part of each review was filled out. Also, these 40 cases are only .133% of the sample of 30,121 reviewed records, and the sum of the weights for these 40 cases is 3,590, which is 0.134% of the sum of the weights (2,671,863) of the 30,121 reviewed records in the sample. We use imputation methods to fill in the missing values because:

i. using a completed data set makes the analysis simpler and less prone to error
ii. imputation methods, adjust for the missing data in a statistically sound and justifiable way.

In this appendix, we show that our final estimates will be insensitive to whatever values we impute for this small amount of missing data. We need to impute values for confidence of causation (question 4.8), degree of disability (question 6.3), severity of negligence (question 8.9), and confidence of negligence (question 9.0). Any combination or all of these questions might be missing on a review. We decided that imputation of a missing item on a review should be based on a highly correlated item that was answered on the same review. The exception was degree of disability, which had a high intra-class correlation coefficient (about .9) and thus was imputed based on the value observed on the other review. Filling in missing items in this way requires the weakest missing data assumptions.

A. Imputation Plan

(Note: One review can have any or all of the following items missing, all of which need to be imputed).

1. Question 4.8, confidence of causation.

   i. Completion of Q4.8 was to be based on the answer to Q9.0 (confidence of negligence); this occurred on 19 reviews.

   ii. if both Q9.0 and Q8.9 were missing, Q4.8 was imputed based on the observed value of Q4.8 on the other review; this occurred on 1 review.

2. Question 6.3, degree of disability.

   Q6.3 was imputed based on the observed value of Q6.3 on the other review; this occurred on 3 reviews.

3. Question 9.0, confidence of negligence.

   i. Completion of Q9.0 was to be based on the answer to Q8.9 (severity of negligence); this occurred on 6 reviews.

   ii. If Q8.9 was missing, Completion of Q9.0 was to be based on the answer to q4.8 (confidence of causation); this occurred on 7 reviews.

4. Question 8.9, severity of negligence.

   i. Completion of Q8.9 was to be based on the answer to Q9.0 (confidence of negligence); this occurred on 14 reviews.

B. Example

As an example, we will consider how to impute a value for a case with question 9.0 (confidence of negligence) answered on a review but question 8.9 (severity of negligence) not answered on the review. We want to impute the value based on the data we have in which both questions 8.9 and 9.0 are answered on a single review (we give each review one half of the case weight), from which we can form the following table of probabilities:

Table 1. Question 8.9 versus question 9.0

Q8.9 (severity of negligence) \ Q9.0 (confidence of negligence)	1	2	3	4	5	6
1	p_{11}	p_{12}	p_{13}	p_{14}	p_{15}	p_{16}
2	p_{21}	p_{22}	p_{23}	p_{24}	p_{25}	p_{26}
3	p_{31}	p_{32}	p_{33}	p_{34}	p_{35}	p_{36}

Table 1 presumably has 18 underlying probabilities:

$\underline{p} = \{p_{ij}\} = \{$ Pr(a reviewer rates Q8.9 an i and Q9.0 a j) $\}$.

C. Multiple Imputation:

In the first step of the imputation task, we need a consistent estimate of $\underline{p}$, say $\hat{\underline{p}}$, and also a consistent estimate of its covariance matrix, say $\hat{Var}(\hat{\underline{p}})$. From the sample, we obtained a weighted ratio estimate, $\hat{\underline{p}}$, using SAS Proc Sesudaan (which is consistent assuming MCAR). However, we could not get an estimate of its covariance matrix using SAS Proc Sesudaan; only an estimate of the standard errors of the elements of $\hat{\underline{p}}$, i.e.,the $\hat{se}(\hat{p}_{ij})$. We did our imputations with the $\hat{Var}(\hat{\underline{p}}) = 0$, because we found that the imputations were insensitive to the estimate of $\hat{Var}(\hat{\underline{p}})$.

We want to transform the elements of $\hat{\underline{p}}$ onto the polychotomous logisitic scale for the imputations:

$$\hat{\theta}_{ij} = \log(\hat{p}_{ij}/\hat{p}_{36}) \quad , \quad ij =/ 36 \ .$$

If we let the vector $\hat{\underline{\theta}} = \{\hat{\theta}_{ij}\}$, then we can obtain $\hat{Var}(\hat{\underline{\theta}})$ using the delta method. We have made this transformation because the elements of $\underline{\theta}$ can take on any value on the whole real line, whereas the elements of $\underline{p}$ must be between 0 and 1 and also must sum to 1.

Step 1:

Draw a value of $\underline{\theta}$ from

$$\underline{\theta} \sim N\left[\hat{\underline{\theta}}, \hat{Var}(\hat{\underline{\theta}})\right] .$$

Step 2:

Transform the drawn value $\underline{\theta}$ back to the probability scale $\underline{p}$ using the inverse of the logit transformation.

Step 3:

Suppose $k = 1,\ldots,n_o$ ($n_o = 40$) cases have item non-response. For example, consider a case with the value of Q9.0 equal to 5 and Q8.9 is missing. We know that this case can only belong to one of the cells p_{15}, p_{25}, or p_{35} in the table of probabilities (i.e., the probabilities in the fifth column of the table). Thus, we can condition on the fifth column of the table. With n_o cases with item non-response, we will have n_o tables of conditional probabilities (although the tables will be the same for individuals with the same pattern of missingness). For case k, we would calculate a set of T_k conditional probabilities $\{p_{1k},\ldots p_{Tk}\}$. For example, for the case discussed above, we have three conditional probabilities:

$$p_{1k} = p_{15}/(p_{15}+p_{25}+p_{35}) \quad ;$$

$$p_{2k} = p_{25}/(p_{15}+p_{25}+p_{35}) \quad ; \quad \text{and}$$

$p_{3k} = p_{35}/(p_{15}+p_{25}+p_{35})$.

Step 4:

Draw n_o independent uniform (0,1) random numbers, say u_k.

Impute, for case k,:

1 if $0 < u_k \leq p_{1k}$;

2 if $p_{1k} < u_k \leq p_{1k} + p_{2k}$;

.
.
.

T if $p_{1k} + \cdots + p_{T-1,k} < u_k \leq 1$.

Imputations for other individuals with missing items will be similar. They will all be based on contingency tables, in which we know an individual can only fall in a certain subset of the cells, where this subset is determined by the value of the question that is observed.

D. Naive imputation (only one imputation for each missing value):

From the sample, we can form a table of weighted cell counts (recall, we give each review one half of the case weight), which corresponds to the probabilities in Table 1.

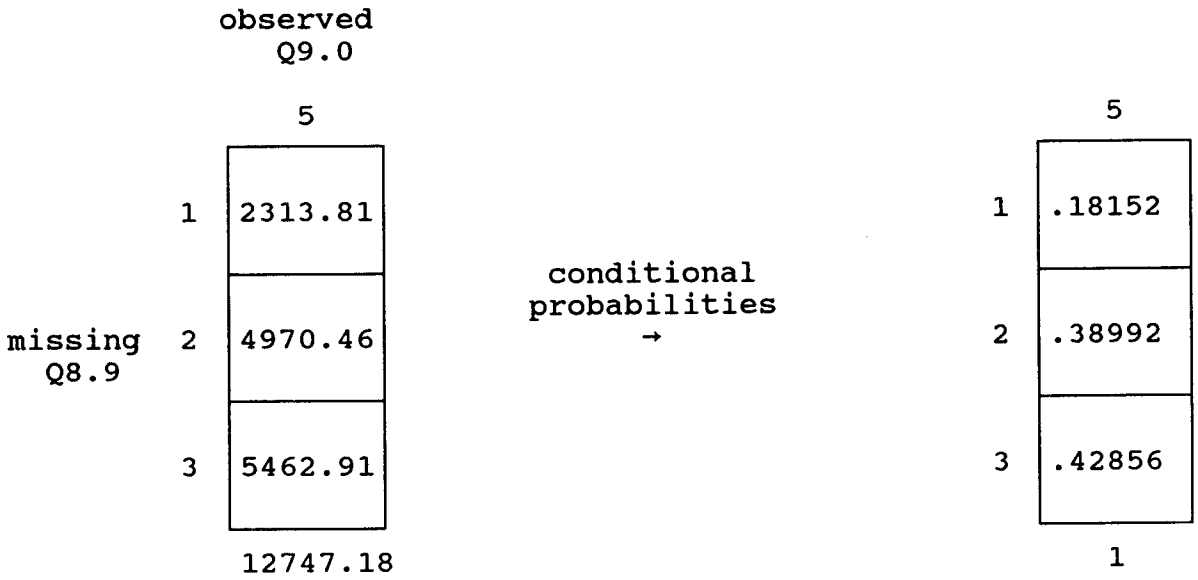
Q9.0 (confidence of negligence)

Q8.9 (severity of negligence)	1	2	3	4	5	6
1	3997.83	7346.32	2725.42	4413.58	2313.81	693.085
2	566.999	2073.13	2754.12	8735.64	4970.46	3004.12
3	560.782	708.202	685.85	2024.28	5482.91	5165.46

Consider the case with Q9.0 equal to 5 and Q8.9 missing.

Step 1:

Since the value of Q9.0 is 5, we know that the case can only belong to one of the cells p_{15}, p_{25}, or p_{35} in the table of probabilities (i.e., the probabilities in the fifth column of the table). Thus, we can condition on the fifth column of the table (since we know the person has to be somewhere in the fifth column):



Step 2:

For the case with Q9.0 equal to 5 and Q8.9 missing, we impute a

1 with probability .18152

2 with probability .38992

3 with probability .42856.

One way to accomplish this is to draw a random number (from the computer) between 0 and 1. If the random number is between 0 and .18152, then impute a 1. If the random number

is between .18152 and .57144 (= .18152 + .38992), then impute a 2. If the random number is between .57144 and 1, then impute a 3.

The random number that was drawn was .591, so we impute a 3 for the missing value of question 8.9 (severity of negli- gence).

It would be statistically sounder, to do M sets of imputations. We will now show, howeever, that a single imputation is sufficient for our purposes.

E. Showing that a single imputation is sufficient

Maximum Imputation Variance

	Negligent AE		AE	
Imputation scheme	$\hat{P}$	S.E.	$\hat{P}$	S.E.
Given imputation	.0101721	.00104691	.0369065	.00244477
Set AE=1, NEGAE=1	.0100414	.00104823	.0364975	.00240109
Set AE=6, NEGAE=6	.0104041	.00104931	.0370364	.00244179

With M imputations,

1) the point estimate is $\bar{P}_M = \sum_{m=1}^{M} \hat{P}_m/M$;

2) the average within-imputation variance is

$$\bar{W}_M = \sum_{m=1}^{M} \hat{V}_m/M \ ;$$

where $\hat{V}_m$ is the estimate of variance for the mth imputation.

3) the between-imputation component is

$$B_M = \sum_{m=1}^{M} (\hat{P}_m - \bar{P}_M)^2/(M-1) \ ; \quad \text{and}$$

4) the estimate of variance is

$$V_M = \bar{W}_M + ((M+1)/M)B_M.$$

Note:

1) The maximum within-imputation variance occurs if the imputed values of an adverse event (AE) are 6 and negative adverse event (NEGAE) are 6 for all imputations. In this case, $B_M = 0$, and

$$V_M = \bar{W}_M = (.00244179)^2 = .0000059623 \quad \text{(for AE)}$$

$$V_M = \bar{W}_M = (.00104931)^2 = .0000011011 \quad \text{(for NEGAE)}$$

2) The maximum between-imputation component occurs when one half (M/2) of the imputations impute values of AE to be 6 and NEGAE to be 6, and one half of the imputations impute values of AE to be 1 and NEGAE to be 1. In this case,

$$\bar{W}_M = (.00240109^2 + .00244179^2)/2 = .0000058638 \quad \text{(for AE)}$$

$$\bar{W}_M = (.00104823^2 + .00104931^2)/2 = .0000010999 \quad \text{(for NEGAE)}$$

and (letting $M \to \infty$),

$$B_M = .5[\hat{P}_1-(\hat{P}_1+\hat{P}_2)/2]^2 + .5[\hat{P}_1-(\hat{P}_1+\hat{P}_2)/2]^2$$

$$= .5[(\hat{P}_1-\hat{P}_2)/2]^2 + .5[(\hat{P}_2-\hat{P}_1)/2]^2 = (P_1-P_2)^2/4 ,$$

where

$\hat{P}_1$ = estimated rate when AE or NEGAE is imputed a '1' ;

$\hat{P}_2$ = estimated rate when AE or NEG is imputed a '6'.

Then,

$$B_M = (.0364975 - .0370364)^2/4 = .000000072603 \quad \text{(for AE)}$$

$$B_M = (.0100414 - .0104041)^2/4 = .000000032888 \quad \text{(for NEGAE)}.$$

If this set of imputations actually occurred,

Negligent AE		AE	
$\hat{P}$	S.E.	$\hat{P}$	S.E.
.010223	.0010643	.036767	.0024365

3) We can estimate the fraction of information about the parameter missing due to nonresponse as $1/(1+\phi)$, where

$\phi = \bar{W}_M/B_M$. For (2),

	Negligent AE	AE
ϕ	33.4446	80.7647
fraction of missing information	.02903	.01223

4) The minimum within imputation variance occurs if the imputed values of AE are 1 and negligent AEs are 1 for all imputations). In this case, $B_M = 0$, and

$$\bar{W}_M = (.00240109)^2 = .0000057652 \quad \text{(for AE)}$$

$$\bar{W}_M = (.00104123)^2 = .0000010988 \quad \text{(for NEGAE)}$$

5) Then, an estimate for the upper bound of the amount of information missing is found by using W_M from (4) and B_M from (2), which gives

	Negligent AE	AE
ϕ	33.4101	79.4073
fraction of missing information	.02906	.01244

These calculations show that because we impute such a small amount of missing data (and the fraction of missing information is so small) that our results are insensitive to the actual imputation used, and also that little will be gained by doing multiple imputations.

ADJUSTING SAMPLE WEIGHTS FOR MISSING AND NON-REVIEWED RECORDS

I. The Adjustment Process

A. Records Missing at Stage 1

The first wave of record reviews left 1,234 records not found either because the hospital could not locate a record or because the record fitting the identifier on the Statewide Planning and Research Cooperative System file did not match the hardcopy record supplied by the hospital. For all of these records, we had available to us information from the SPARCS about patient clinical and demographic characteristics.

Where data internal to the sample are available, the common method of adjustment for missing records is the sample weighting adjustment,[1] by which one (1) divides the sample into weighting adjustment subclasses, and (2) adjusts the weights of the records in each subclass by the inverse of the probability of a record in that subclass being located. For this weighting adjustment, we needed to determine the subclasses to be used for the adjustment and the weighted proportion of cases located for each of these subclasses.

1. **Subclasses for Missing Record Adjustment.** Our goal was to ensure that the missing records did not distort the incidence of adverse events. To determine appropriate subclasses for missing record adjustment, we examined the AE rates and the missing record rates for a variety of patient characteristics (subclasses). Our analysis suggested that adverse events and missing records varied with the hospital, the sampling weight group (base weight, elderly, low-risk newborn, high risk newborn, and high-risk surgery), age, death on discharge, and Medicare

[1] Kalton, Kasprzyk, The treatment of missing survey data, 12 Survey Methodology 1-16 (1986)

reimbursement. We found little difference in adverse event and missing record rates by patient sex. We also knew that old age, death on discharge, and Medicare status were all characteristics of the sampling weight group "elderly." For these reasons, we chose to base weighting adjustment subclasses on hospital and on four weight groups (base, elder, low risk obstetrics, and all high risk). With 46 hospitals with missing records and 4 possible weight groups per hospital, this method produced 184 potential weighting adjustment subclasses. Eleven of these subclasses contained no observations; a number of them contained few observations.

2. **Missing Adjustment Within Subclasses.** Where all adjustment subclasses contain sufficient numbers of non-missing records, the standard method for missing record adjustment would be to increase the weights of the non-missing records by multiplying each record's sampling weight by the inverse of the probability of finding a record in the adjustment subclass. If wt_{ijk} is the sampling weight for the kth observation in adjustment subclass ij, and $m_{ijk}=1$, if the observation is non-missing, and $=0$, if it is missing, then the probability of being located is simply:

$$p_{ijk}=(\sum_{k}(wt_{ijk}*m_{ijk})/\sum_{k}(wt_{ijk}).$$

Stated simply, the probability of a record being located is the sum of the weights of the records located in a sample subclass divided by the total records in the sample subclass, where the subclass if formed by the weighted counts of observations cross classified by hospitals i and weight group j. The weight adjusted for any observation then becomes:

$$adjwt_{ijk}=m_{ijk}*wt_{ijk}*\sum_{k}(wt_{ijk})/(\sum_{k}(wt_{ijk}*m_{ijk}).$$

Here, $adjwt_{ijk}=0$ if the observation is missing and is the original weight multiplied by the inverse of the probability of being located in the adjustment subclass if the record was located.

We found this method of adjustment inappropriate for our data because for some adjustment cells, p_{ijk} was low, and thus the weighting subclasses were unstable. This instability produced a significant increase in weights for located records in some of the adjustment subclasses. To avoid over-adjustment from small subclasses, we chose to smooth the adjustment by raking ratio estimation.

Using the raking method, the marginal distributions of the observations with new weights equal the marginal distributions of all records in the sample with respect to the hospital and sampling weight group variables (base weight, elderly, low-risk obstetrics, high-risk).

Raking is the classical method of adjusting cross-classified tables for new marginal totals.[2] It corresponds to iterative proportional fitting of contingency tables. Applying this technique to the problem at hand, we arranged the located records and all the sample records in two matrices of 184 cross-classified cells formed by 46 hospitals and 4 weighting groups.

Next, we used the iterative proportional fitting (IPF) function of PROC IML[3] to scale the total weights of cases located in each adjustment cell to satisfy the constraints of the row and column margins of total weights of all the sample cases.[4] Thus, if fit_{ij}=the scaled up total weight of located records in the

---

[2] Y.M.M. Bishop, S.E. Fienberg, P.W. Holland, Discrete multivariate analysis: theory and practice. Cambridge, MA: MIT Press (1975)

[3] L.D. Crum, ed. SAS/IML User's Guide, v. 5. Cary, NC: SAS Institute, 1986.

[4] R.J.A. Little, D.B. Rubin, Statistical analysis with missing data. New York: Wiley, 1987:58-60.

cell defined by hospital i and group j, and N_{ij}= the total weight of all sample records defined by these variables, raking by IPF will satisfy the constraints:

$$fit_{i+} = \sum_{j} fit_{ij} = N_{i+}, \quad \text{and}$$

$$fit_{+j} = \sum_{i} fit_{ij} = N_{+j}.$$

In terms of the IPF algorithm, the data matrix is the marginal totals of the weighted counts of both missing and located records classified by hospital and by weight group. The start matrix is the cross-classified matrix of weighted counts of located records. Using this method,

$$adjwt_{ijk} = m_{ijk} * wt_{ijk} * (fit_{ij}) / (\Sigma_{k}(wt_{ijk} * m_{ijk}),$$

where $\Sigma(fit_{ij})$=the fitted value after fitting an independence model for the weighted cell counts for hospital i and weight group j.

B. Records Referred, Not reviewed

After first adjusting sampling weights for those records not located for the Stage 1 screen, we performed a similar adjustment for those cases referred from Stage 1 for physician review, but not actually reviewed. We chose adjustment subclasses and then increased the weights of the 7,743 records referred and reviewed to represent all 7,817 cases referred.

1. Subclasses for Cases Referred but not Reviewed.

Weighting adjustment subclasses for these cases were formed in the same manner as those for the missing cases. At twenty-six hospitals, the MRAs referred records for physician review but

could not obtain those reviews. Table 2 lists the frequency of cases by hospital.

Table 2

Number of hospitals	Number of Cases Referred, Not Reviewed/	Total Cases Referred, Not Reviewed
1	21	21
3	6	18
1	4	7
3	3	9
4	2	8
14	1	14
26		74

2. Adjustment within subclasses. As with the missing cases, the problem of small adjustment cells suggested that a reweighting by cell would give a few observations large increases in weights. Therefore, we chose to smooth the cell weight adjustment, again using raking ratios. Adjustments were confined to hospitals with records not reviewed. The start matrix was the total weights of records referred and reviewed cross classified by hospital and weighting group; the data matrix was the total weights of all records referred.

II. Results of Adjustment Process

A. Records Missing at Stage 1

As a result of the adjustment to weighting, the weights of 1,234 records were set to 0. Although they remain in the sample, their weights are assigned to the non-missing cases. For the remaining 30,195 cases the sampling weights calculated from the probability of selecting the record (weight= 1/p, where p=probability of selection) were adjusted by factors from 0.986 to 1.27 (i.e., weights were decreased by 1.4% or less and increased by 27% or less). Only 3.6% of these cases (1,092) needed an increase in weight of 20% or more, and 75% (22,711) of the cases reweighted were changed by less than 5%.

B. Records Referred, Not Reviewed

Of the 30,195 records not missing at stage 1, 83.2% (25,132) remained unchanged by this weighting adjustment. The 74 records not reviewed were adjusted to 0 sampling weights. Adjustments for the 4,989 remaining records ranged from a decrease of 0.2% to an increase of 11.2%. Only 7% (350) of the remaining cases needed an adjustment of more than 5%.

C. Comparisons with Populations

Tables 3 through 6 compare distributions of the weighted sample of records located and reviewed before and after the adjustments described above. Column (1) gives the distributions of the 30,121 records located and reviewed using the weights computed from the probability of selecting the record from the population of all discharges in 1984 (see Technical Appendix 4.II.2). Column (2) shows the same discharges with weights modified to account for 1,234 missing records and 74 records not reviewed. Column (3) gives the fraction of records that were located and reviewed.

Table 3

AGE	Weighted % Before Adjustment	Weighted % After Adjustment	Records Located + Reviewed as wtd % of Sample
Newborn	9.6	9.5	97.1
0-15	8.5	8.5	96.4
16-44	34.4	34.3	96.3
45-69	26.6	26.5	95.0
70+	20.9	21.1	94.9

Table 4

RACE	Weighted % Before Adjustment	Weighted % After Adjustment	Records Located + Reviewed as wtd % of Sample
Black	14.9	15.4	92.1
Hispanic	5.3	5.4	94.9
White/Other	79.8	79.3	96.5

Table 5

PRIMARY REIMBURSEMENT	Weighted % Before Adjustment	Weighted % After Adjustment	Records Located + Reviewed as wtd % of Sample
Self pay	7.1	7.1	96.9
Medicare	28.6	28.7	94.5
Medicaid	13.2	13.5	93.0
BC/Others	51.2	50.7	97.0

Table 6

DIAGNOSIS GROUP[5] AT INDEX HOSPITALIZATION	Weighted % Before Adjustment	Weighted % After Adjustment	Records Located + Reviewed as wtd % of Sample
General Med, Others	5.0	5.1	96.2
Cardiology	10.2	10.3	95.2
Dentistry	0.7	0.7	98.5
Dermatology	0.4	0.4	95.5
Endocrinology	1.8	1.8	92.8
Gastroenterology	5.8	5.8	95.3
General Surgery	11.8	11.9	95.8
Gynecology	5.7	5.7	95.6
Hematology	0.8	0.8	93.3
Neonatology	9.9	9.7	97.2
Nephrology	1.7	1.7	93.3
Neurology	4.4	4.4	95.3
Neurosurgery	1.1	1.1	96.2
Obstetrics	12.7	12.6	96.5
Oncology	3.8	3.8	93.8
Ophthalmology	2.7	2.7	97.8
Orthopedic Surgery	6.2	6.2	96.2
Otolaryngology	3.3	3.3	97.6
Pulmonary Disease	5.8	5.9	94.8
Rheumatology	0.6	0.6	92.9
Cardiac/Thoracic Sur	1.2	1.2	96.0
Urology	3.7	3.7	94.9
Vascular Surgery	0.7	0.7	91.9

[5] Diagnosis groups are formed by grouping patients by the DRG listed in the SPARCS dataset for the sampled hospitalization.

RATIO ESTIMATES AND VARIANCE ESTIMATES FOR IMPLICITLY STRATIFIED UNEQUAL CLUSTER SAMPLING USING SAS PROC SESUDAAN

I. The Ratio Estimator

Our sample consists of 31 clusters. The sampling design is one of unequal cluster sampling with implicit stratification of clusters by systematic sampling. This appendix gives the formulas we used for computing weighted rates and their sampling errors. In calculating these estimates, we needed to decide which stratification plan was most similar to our design (systematic sampling with implicit stratification). The sampling frame naturally formed H = 7 strata based on geographic, teaching, and ownership characteristics of the hospitals:

1) MSA, nonprofit, nonteaching, downstate (Westchester or below)
2) MSA, nonprofit, nonteaching, upstate (above Westchester)
3) MSA, nonprofit, teaching, upstate
4) MSA, nonprofit, teaching, downstate
5) MSA, for profit hospitals
6) All Non-MSA hospitals, regardless of ownership
7) MSA, government hospitals

We used these 7 strata in calculating the estimates.

In the h^{th} strata, we sampled n_h out of the N_h PSUs (i.e., clusters, which are hospitals). The sample and population strata sizes are given in Table 1.

Table 1. Sample and population cluster sizes

h (strata)	n_h	N_h
1	5	30
2	5	36
3	4	13
4	9	31
5	2	11
6	2	22
7	4	19

Further, for the i^{th} PSU in stratum h, we sampled m_{hi} out of the M_{hi} units. In this memo, we consider the binary response variable AE (0/1). Then, we let

$$Y_{hij} = \begin{cases} 1 & \text{if the } j^{th} \text{ individual in the } i^{th} \text{ PSU within the } h^{th} \text{ stratum is an AE} \\ 0 & \text{otherwise} \end{cases}$$

Finally, the hij^{th} individual has weight W_{hij}. Then , the ratio estimate of the probability of having an AE, $\hat{R}$, is

$$\hat{R} = \left[\sum_{h=1}^{H} \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} W_{hij} Y_{hij} \right] / \left[\sum_{h=1}^{H} \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} W_{hij} \right] . \qquad (1)$$

II. Variance Estimator of Ratio Estimator

If we let

$$y = \left[\sum_{h=1}^{H} \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} W_{hij} Y_{hij} \right]$$

and

$$x = \left[\sum_{h=1}^{H} \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} W_{hij} \right] \quad ;$$

then, using a first-order Taylor series expansion, the large sample variance of $\hat{R}$ is

$$\mathrm{Var}(\hat{R}) = [\mathrm{Var}(y) + R^2 \cdot \mathrm{Var}(x) - 2R \cdot \mathrm{Cov}(x,y)]/X^2, \qquad (2)$$

where R is the true proportion of AEs in the population and X is the total number of individuals in the population.

Alternatively, if we let

$$Z = [y - Rx]/X = [\hat{R} - R](x/X) \qquad (3)$$

and

$$Z_{hij} = W_{hij}[Y_{hij} - R]/X, \qquad (4)$$

then

$$\mathrm{Var}(\hat{R}) = \mathrm{Var}(Z) = \mathrm{Var}\left[\sum_{h=1}^{H} \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} Z_{hij} \right] . \qquad (5)$$

We estimate $\mathrm{Var}(\hat{R})$ by substituting the estimates of R ($\hat{R}$) and X (x) in (5). (We denote (4) with $\hat{R}$ substituted for R and x substituted for X by z_{hij}.) Then, including the finite sample population correction factors in the estimates of Var(y), Var(x), and Cov(x,y), the Taylor variance approximation (used by Proc Sesudaan) for two-stage stratified sampling is

$$\hat{\mathrm{Var}}(\hat{R}) = \sum_{h=1}^{H} (1-F_h) n_h s_h^2 + \sum_{h=1}^{H} F_h \sum_{i=1}^{n_h} (1-f_{hi}) m_{hi} s_{hi}^2, \qquad (6)$$

where

$$s_h^2 = \sum_{i=1}^{n_h} [z_{hi+} - \bar{z}_{h.+}]^2/(n_h - 1) \qquad (7)$$

is the variability in the clusters within the h^{th} stratum, with

$$z_{hi+} = \sum_{j=1}^{m_{hi}} z_{hij} \quad \text{(the cluster sum)},$$

and

$$\bar{z}_{h.+} = \sum_{i=1}^{n_h} z_{hi+}/n_h \quad \text{(the stratum mean of the cluster sums)}.$$

Also, s_{hi}^2 is the variablity of the individuals within cluster i in stratum h:

$$s_{hi}^2 = \sum_{j=1}^{m_{hi}} [z_{hij} - \bar{z}_{hi+}]^2/(m_{hi} - 1), \qquad (8)$$

where

$$\bar{z}_{hi+} = \sum_{j=1}^{m_{hi}} z_{hij}/m_{hi} \quad \text{(the mean of } i^{th} \text{ cluster in stratum h)}.$$

The quantities F_h and f_{hi} are the first- and second-stage sampling rates, respectively, that is,

$$F_h = n_h/N_h \text{ and } f_{hi} = m_{hi}/M_{hi}.$$

Suppose that we are interested in the covariance between two rates, $\hat{R}_k$ and $\hat{R}_m$, for different response variables (say, AE and negligent AE). Following similar reasoning as in (6),

$$\hat{Cov}(\hat{R}_k,\hat{R}_m) = \sum_{h=1}^{H} (1-F_h) n_h s_h^{km} + \sum_{h=1}^{H} F_h \sum_{i=1}^{n_h} (1-f_{hi}) m_{hi} s_{hi}^{km}, \qquad (9)$$

where

$$s_h^{km} = \sum_{i=1}^{n_h} [z_{hi+} - \bar{z}_{h.+}][u_{hi+} - \bar{u}_{h.+}]/(n_h - 1) \qquad (10)$$

and

$$s_{hi}^{km} = \sum_{j=1}^{m_{hi}} [z_{hij} - \bar{z}_{hi+}][u_{hij} - \bar{u}_{hi+}]/(m_{hi} - 1), \qquad (11)$$

In (10) and (11), z_{hij} corresponds to variable 1 (e.g., AE) and u_{hij} corresponds to variable 2 (e.g., NEGAE). Note also that $\hat{R}_k$ and $\hat{R}_m$ also could correspond to the same response variable (e.g., AE) for two subgroups of the population (e.g., male and female). In this case

$$\hat{R}_k = \left[\sum_{h=1}^{H} \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} W_{hij} Y_{hij} k_{hij} \right] / \left[\sum_{h=1}^{H} \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} W_{hij} k_{hij} \right] , \qquad (12)$$

where

$$k_{hij} = \begin{cases} 1 & \text{if individual hij is in subgroup k} \\ 0 & \text{otherwise} \end{cases}$$

Because of the cluster design, $\hat{R}_k$ and $\hat{R}_m$ may be correlated since individuals from the same hospital (i.e., cluster) contribute to both $\hat{R}_k$ and $\hat{R}_m$. Then,

$$z_{hij} = \{W_{hij}[Y_{hij} - R_k]/\hat{x}_k\}k_{hij} , \qquad (13)$$

and

$$x_k = \left[\sum_{h=1}^{H} \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} W_{hij} k_{hij} \right],$$

Also, $u_{hij} = \{W_{hij}[Y_{hij} - R_m]/\hat{x}_m\}m_{hij}$ and x_m are defined similarly. Equations (6) and (9) are still used to calculate the variance and covariance, respectively.

$\hat{Cov}(\hat{R}_k,\hat{R}_m)$ is not directly computed in Proc Sesudaan, however. Proc Sesudaan computes only the variance of the difference of the ratio estimates: that is, $\hat{Var}(\hat{R}_k-\hat{R}_m)$. From probability theory,

$$\hat{Var}(\hat{R}_k-\hat{R}_m) = \hat{Var}(R_k) + \hat{Var}(R_m) - 2\hat{Cov}(\hat{R}_k,R_m) \qquad (14)$$

Then, from (14),

$$\hat{Cov}(\hat{R}_k,\hat{R}_m) = .5\{ \hat{Var}(\hat{R}_k) + \hat{Var}(\hat{R}_m) - \hat{Var}(\hat{R}_k - \hat{R}_m) \}. \qquad (15)$$

The 3 components on the right hand side of (15) are all computed in Sesudaan.

III. Wald Test

Suppose that we are interested in testing whether D rates are equal to each other (D refers to the rates, say AE, for D subgroups of a variable, say race, with subgroups black, white/others, and Hispanic); The null hypothesis is

$$H_o: R_1 = R_2 = \cdots = R_D .$$

We can also write this null hypothesis as $C\underline{R} = \underline{0}$, for the appropriate contrast matrix C and where $\underline{R} = [R_1,\ldots,R_D]'$.
We can use the $\hat{\underline{R}} = (\hat{R}_k)$ to test this null hypothesis. Because of the clustering, $\hat{R}_k$ and $\hat{R}_m$ are correlated. Our estimate of the covariance matrix, $\hat{Var}(\hat{\underline{R}})$, is obtained using (15). We use as our test statistic

$$G^2 = (C\hat{\underline{R}})'[C\hat{Var}(\hat{\underline{R}})C'](C\hat{\underline{R}}) ,$$

which, under the null hypothesis of the equality of rates, is approximately chi-square with (D-1) degrees of freedom. Instead of using an F statistic with (D-1) numerator degrees of freedom, and df_2 denominator degrees of freedom, we use the chi-square statistic. This is because the degrees of freedom (which is df_2)

associated with the variance estimate (given in (6)) is approximately 7,000 (approximately the sum of the degrees of freedom associated with the two stages) and an F-statistic with a large df_2 is approximately a chi-square statistic. The first stage variance (s_h^2) has 24 degrees of freedom (31 clusters - 7 strata) and the second-stage variance (s_{hi}^2) has roughly 6,993 degrees of freedom, which is calculated using the approximation

df of s_{hi}^2 = (average cluster size minus 1) x (number of strata),

= [(31000/31) - 1] x 7

= [1000 - 1] x 7 = 6993 .

Although the degrees of freedom of s_{hi}^2 is calculated using a rough approximation, df_2 is so large that the chi-square approximation is justified.

When making a two-group comparison, in which we use the statistic $(\hat{R}_k - \hat{R}_m)/\{\hat{Var}(\hat{R}_k) + \hat{Var}(\hat{R}_m) - 2\hat{Cov}(\hat{R}_k,\hat{R}_m)\}^{1/2}$, we use a normal approximation instead of a t statistic (with 7,000 degrees of freedom) for the same reason we gave for using the chi-square approximation.

III. Variance Approximation

In some cases, it was necessary to obtain an estimate of $\hat{Var}(\hat{R})$ with the first-stage sampling variability; that is,

$$\hat{Var}(\hat{R}) = (1-f) \sum_{h=1}^{H} n_h s_h^2 , \qquad (16)$$

where $f = F_h f_{hi}$ is the overall sampling rate and is approximately 1/86 for all individuals. Kalton (1988) shows that (16) is an appropriate (i.e., consistent) estimate and also requires fewer assumptions about the sampling design associated with the second and latter stages (since it only uses the variability associated with the first stage). Unlike the two-stage estimate of variance, however, (16) only has 24 degrees of freedom (since we are only using the first-stage variance). When calculating AE and negligent AE rates associated with disability, we treated the two reviews on the same individual as separate measurements within an individual (thus, we can think of an individual as another level of clustering). Instead of using (6) for estimating the variance of these rates, (16) was used since it does not require assumptions about sampling within an individual. For all other variance estimates, (6) was used. To implement formula (16) with Proc Sesudaan (which uses formula (6)), we let $F_h = f = 1/86$ for all h, or, equivalently, $N_h = n_h/f = 86n_h$. Also, we let $m_{hi} = M_{hi}$ for all hi so that $f_{hi} = 1$ for all hi, and thus the second term in (6) is 0.

Classifying Patients into DRG Groups

Many of the analyses in the Study require comparisons of adverse event rates among groups of patients that have different proportions of diagnoses. In order to ensure that these comparisons involve similar diagnostic outcomes, we standardized by Diagnosis Related Groups (DRG).

The SPARCS data base assigns each patient discharge to one of 470 DRGs. Major Diagnostic Codes (MDC) are available to collapse the 470 DRGs into 23 groups of related diagnoses. Unfortunately, both these classifications would provide grouping with too small numbers for stable rates. Therefore, in order to obtain broader groups, a list of all 470 DRGs was distributed to three senior physicians who were asked to rate on a scale of 1 to 6 the likelihood that a patient in that DRG would experience an AE. A score of 6, for example, would mean that patients grouped in the DRG would experience the highest risk of an AE.

All DRGs received at least one AE likelihood rating. The scores of the three physicians were averaged, and the distribution is plotted below. By selecting natural break points in the distribution, the scores were grouped into four risk categories.

The risk groups formed by these physician judgments were validated first by comparing the AE rates among these groups using the data from the Medical Practice Study pilot project, completed in April 1987. They were validated again with the 30,121 observations of this Study. The rate of adverse events by DRG Level is shown in Table 5.6 in Chapter 5. Both data sets exhibited monotonic increases in AE rates with DRG level.

TABLE 1

Average DRG Score as Rated by 3 MDs

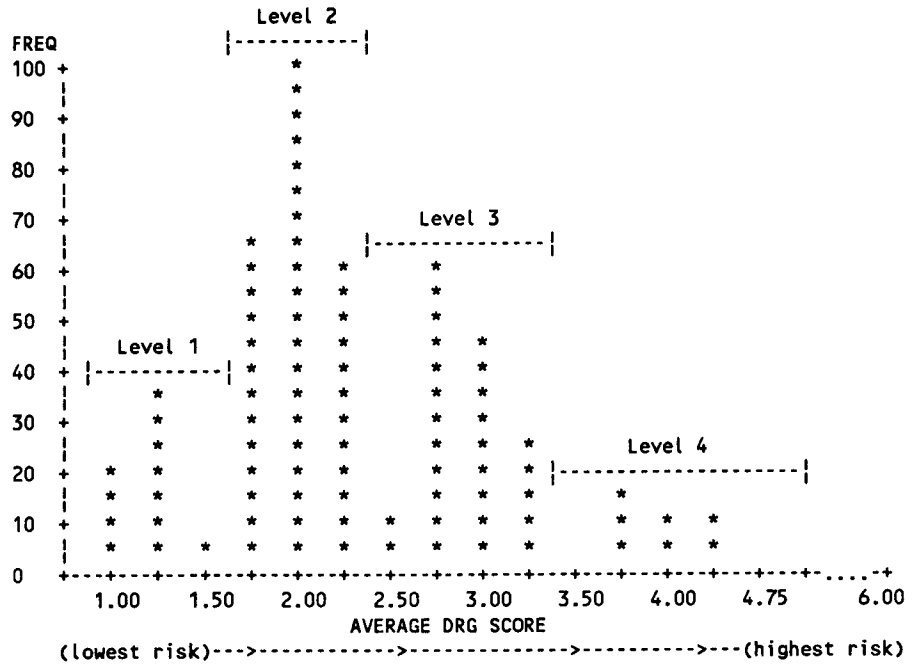


TABLE 2

DRGs by DRG Level

DRG LEVEL 1

006 Carpal Tunnel Release
012 Degenerative Nervous System Disorders
016 Nonspecific Cerebrovascular Disorders with C.C.
017 Nonspecific Cerebrovascular Disorders w/o C.C.
018 Cranial & Peripheral Nerve Disorders Age GE 70 and/or C.C.
019 Cranial & Peripheral Nerve Disorders Age LT 70 w/o C.C.
024 Seizure & Headache Age 18-69 w/o C.C.
061 Myringotomy Age GE 18
062 Myringotomy Age 0-17
065 Dysequilibrium
163 Hernia Procedures Age 0-17
168 Procedures on the Mouth Age GE 70 and/or C.C.
169 Procedures on the Mouth Age LT 70 w/o C.C.
185 Dental & Oral Dis. Exc Extractions & Restorations, Age GE 18
186 Dental & Oral Dis. Exc Extractions & Restorations, Age 0-17
187 Dental Extractions & Restorations
228 Ganglion (Hand) Procedures
237 Sprains, Strains & Dislocations of Hip, Pelvis & Thigh
280 Trauma to the Skin, Subcut Tiss & Breast Age GE 70 &/or C.C.
281 Trauma to the Skin, Subcut Tiss & Breast Age 18-69 w/o C.C.
282 Trauma to the Skin, Subcut Tiss & Breast Age 0-17
283 Minor Skin Disorders Age GE 70 and/or C.C.
284 Minor Skin Disorders Age LT 70 w/o C.C.
369 Menstrual & Other Female Reproductive System Disorders
382 False Labor
383 Other Antepartum Diagnoses with Medical Complications
384 Other Antepartum Diagnoses w/o Medical Complications
391 Normal Newborns
395 Red Blood Cell Disorders Age GE 18
396 Red Blood Cell Disorders Age 0-17
420 Fever of Unknown Origin Age 18-69 w/o C.C.
421 Viral Illness Age GE 18
422 Viral Illness & Fever of Unknown Origin Age 0-17
423 Other Infectious & Parasitic Diseases Diagnoses
425 Acute Adjust React & Disturbances of Psychosocial Dysfunction
426 Depressive Neuroses
427 Neuroses Except Depressive
428 Disorders of Personality & Impulse Control
429 Organic Disturbances & Mental Retardation
430 Psychoses
431 Childhood Mental Disorders
432 Other Diagnoses of Mental Disorders
433 Substance Use & Subst Induced Organic Mental Disorder, Left AMA
435 Drug Use Except Dependence

437 Alcohol Use Except Dependence
438 Alcohol & Substance Induced Organic Mental Syndrome
447 Allergic Reactions Age GE 18
448 Allergic Reactions Age 0-17
461 O.R. Proc with Diagnoses of Other Contact with Health Services
462 Rehabilitation
463 Signs & Symptoms with C.C.
464 Signs & Symptoms w/o C.C.
465 Aftercare with History of Malignancy as Secondary Dx
466 Aftercare w/o History of Malignancy as Secondary Dx
467 Other Factors Influencing Health Status
468 Operating Procedure Unrelated to Diagnosis
469 Valid Diagnosis Unacceptable as Principal Diagnosis
470 Discharge with Invalid Data

DRG LEVEL 2

007 Periph & Cranial Nerve & Other Nerve Syst Proc Age GE 70 &/or C.C.
008 Periph & Cranial Nerve & Other Nerve Syst Proc Age LT 70 w/o C.C.
013 Multiple Sclerosis & Cerebellar Ataxia
015 Transient Ischemic Attacks
020 Nervous System Infection Except Viral Meningitis
021 Viral Meningitis
022 Hypertensive Encephalopathy
023 Nontraumatic Stupor & Coma
025 Seizure & Headache Age 18-69 w/o C.C.
026 Seizure & Headache Age 0-17
031 Concussion Age GE 70 and/or C.C.
032 Concussion Age 18-69 w/o C.C.
033 Concussion Age 0-17
034 Other Disorders of Nervous System Age GE 70 and/or C.C.
035 Other Disorders of Nervous System Age LT 70 w/o C.C.
036 Retinal Procedures
037 Orbital Procedures
050 Sialoadenectomy
051 Salivary Gland Procedures Except Sialoadenectomy
052 Cleft Lip & Palate Repair
053 Sinus & Mastoid Procedures Age GE 18
054 Sinus & Mastoid Procedures Age 0-17
055 Miscellaneous Ear, Nose & Throat Procedures
056 Rhinoplasty
057 T&A Proc Except Tonsillectomy &/or Adenoidectomy Age GE 18
058 T&A Proc Except Tonsillectomy &/or Adenoidectomy Age 0-17
059 Tonsillectomy and/or Adenoidectomy Age GE 18
060 Tonsillectomy and/or Adenoidectomy Age 0-17
063 Other Ear, Nose & Throat O.R. Procedures
066 Epistaxis
067 Epiglottitis
068 Otitis Media & URI Age GE 70 and/or C.C.

069 Otitis Media & URI Age 18-69 w/o C.C.
070 Otitis Media & URI Age 0-17
071 Laryngotracheitis
072 Nasal Trauma & Deformity
073 Other Ear, Nose & Throat Diagnoses Age GE 18
074 Other Ear, Nose & Throat Diagnoses Age 0-17
081 Respiratory Infections & Inflammations Age 0-17
090 Simple Pneumonia & Pleurisy Age 18-69 w/o C.C.
091 Simple Pneumonia & Pleurisy Age 0-17
093 Interstitial Lung Disease Age LT 70 w/o C.C.
096 Bronchitis & Asthma GE 70 and/or C.C.
097 Bronchitis & Asthma Age 18-69 w/o C.C.
098 Bronchitis & Asthma Age 0-17
099 Respiratory Signs & Symptoms Age GE 70 and/or C.C.
100 Respiratory Signs & Symptoms Age LT 70 w/o C.C.
101 Other Respiratory Diagnoses Age GE 70 and/or C.C.
102 Other Respiratory Diagnoses Age LT 70
117 Cardiac Pacemaker Replace & Revis Exc Pulse Gen Repl Only
118 Cardiac Pacemaker Pulse Generator Replacement Only
119 Vein Ligation & Stripping
128 Deep Vein Thrombophlebitis
133 Atherosclerosis Age LT w/o C.C.
134 Hypertension
140 Angina Pectoris
144 Other Circulatory Diagnoses with C.C.
145 Other Circulatory Diagnoses w/o C.C.
151 Peritoneal Adhesiolysis Age LT 70 w/o C.C.
153 Minor Small & Large Bowel Procedures Age LT 70 w/o C.C.
158 Anal Procedures Age LT 70 w/o C.C.
159 Hernia Procedures Except Inguinal & Femoral Age GE 70 and/or C.C.
160 Hernia Procedures Except Inguinal & Femoral Age 18-69 w/o C.C.
161 Inguinal & Femoral Hernia Procedures Age GE 70 and/or C.C.
162 Inguinal & Femoral Hernia Procedures Age 18-69 w/o C.C.
165 Appendectomy with Complicated Princ. Diag Age LT 70 w/o C.C.
166 Appendectomy with Complicated Princ. Diag Age GE 70 and/or C.C.
167 Appendectomy w/o Complicated Princ. Diag Age LT 70 w/o C.C.
170 Other Digestive System Procedures Age GE 70 and/or C.C.
171 Other Digestive System Procedures Age LT 70 w/o C.C.
177 Uncomplicated Peptic Ulcer GE 70 and/or C.C.
178 Uncomplicated Peptic Ulcer LT 70 w/o C.C.
182 Esophagitis, Gastroent. & Misc. Digest. Dis Age GE 70 &/or C.C.
183 Esophagitis, Gastroent. & Misc. Digest. Dis Age 18-69 w/o C.C.
184 Esophagitis, Gastroent. & Misc. Digest. Disorders Age 0-17
188 Other Digestive System Diagnoses Age Ge 70 and/or C.C.
189 Other Digestive System Diagnoses Age 18-69 w/o C.C.
190 Other Digestive System Diagnoses Age 0-17
216 Biopsies of Musculoskeletal System & Connective Tissue
217 Wnd Debrid & Skn Grft Exc Hand, For Muscskeletal & Conn.Tiss.Dis.
225 Foot Procedures
226 Soft Tissue Procedures Age GE 70 and/or C.C.
227 Soft Tissue Procedures Age LT 70 w/o C.C.

229 Hand Procedures Except Ganglion
230 Local Excision & Removal of Int Fix Devices of Hip & Femur
231 Local Excision & Removal of Int Fix Devices Except Hip & Femur
232 Arthroscopy
233 Other Musculoskelet Sys & Conn Tiss O.R. Proc Age GE 70 &/or C.C.
234 Other Musculoskelet Sys & Conn Tiss O.R. Proc Age LT 70 w/o C.C.
235 Fractures of Femur
236 Fractures of Hip & Pelvis
238 Osteomyelitis
239 Pathological Fractures & Musculoskeletal & Conn. Tiss. Malignancy
240 Connective Tissue Disorders Age GE 70 and/or C.C.
241 Connective Tissue Disorders Age LT 70 w/o C.C.
242 Septic Arthritis
243 Medical Back Problems
244 Bone Diseases & Septic Arthropathy Age GE 70 and/or C.C.
245 Bone Diseases & Septic Arthropathy Age LT 70 w/o C.C.
246 Non-specific Arthropathies
247 Signs & Symtoms of Musculoskeletal System & Conn Tissue
248 Tendonitis, Myositis & Bursitis
249 Aftercare, Musculoskeletal System & Connective Tissue
250 Fx,Sprns,Strns & Disl of Forearm,Hand,Foot Age GE 70 &/or C.C.
251 Fx,Sprns,Strns & Disl of Forearm,Hand,Foot Age 18-69 w/o C.C.
252 Fx,Sprns,Strns & Disl of Forearm,Hand,Foot Age 0-17
253 Fx,Sprns,Strns & Disl of Uparm,Lowleg Ex Foot Age GE 70 &/or C.C.
254 Fx,Sprns,Strns & Disl of Uparm,Lowleg Ex Foot Age 18-69 w/o C.C.
255 Fx,Sprns,Strns & Disl of Uparm,Lowleg Ex Foot Age 0-17
256 Other Diagnoses of Musculoskeletal System & Connective Tissue
261 Breast Proc for Non-Malig Except Biopsy & Loc Exc
262 Breast Biopsy & Local Excision for Non-Malignancy
263 Skin Grafts for Skin Ulcer or Cellulitis Age GE 70 and/or C.C.
264 Skin Grafts for Skin Ulcer or Cellulitis Age LT 70 w/o C.C.
265 Skin Grafts Except for Skin Ulcer or Cellulitis with C.C.
266 Skin Grafts Except for Skin Ulcer or Cellulitis w/o C.C.
267 Perianal & Pilonidal Procedures
268 Skin, Subcutaneous Tissue & Breast Plastic Procedures
269 Other Skin, Subcut Tiss & Breast O.R. Proc Age GE 70 &/or C.C.
270 Other Skin, Subcut Tiss & Breast O.R. Proc Age LT 70 w/o C.C.
271 Skin Ulcers
272 Major Skin Disorders Age GE 70 and/or C.C.
273 Major Skin Disorders Age LT 70 w/o C.C.
276 Non-malignant Breast Disorders
277 Cellulitis Age GE 70 and/or C.C.
278 Cellulitis Age 18-69 w/o C.C.
279 Cellulitis Age 0-17
287 Skin Grafts & Wound Debride for Endoc, Nutrit & Metab Disorders
288 O.R. Procedures for Obesity
289 Parathyroid Procedures
291 Thyroglossal Procedures
292 Other Endocrine, Nutrit & Metab O.R. Proc Age GE 70 and/or C.C.
293 Other Endocrine, Nutrit & Metab O.R. Proc Age LT 70 w/o C.C.
294 Diabetes Age GE 36

295 Diabetes Age 0-35
296 Nutritional & Misc. Metabolic Disorders Age GE 70 and/or C.C.
297 Nutritional & Misc. Metabolic Disorders Age 18-69 w/o C.C.
298 Nutritional & Misc. Metabolic Disorders Age 0-17
299 Inborn Errors of Metabolism
300 Endocrine Disorders Age GE 70 and/or C.C.
301 Endocrine Disorders Age LT 70 w/o C.C.
308 Minor Bladder Procedures Age GE 70 and/or C.C.
309 Minor Bladder Procedures Age LT 70 w/o C.C.
311 Transurethral Procedures Age LT 70 w/o C.C.
312 Urethral Procedrues, Age GE 70 and/or C.C.
313 Urethral Procedures, Age 18-69 w/o C.C.
314 Urethral Procedures, Age 0-17
316 Renal Failure w/o Dialysis
317 Renal Failure with Dialysis
318 Kidney & Urinary Tract Neoplasms Age GE 70 and/or C.C.
319 Kidney & Urinary Tract Neoplasms Age LT 70 w/o C.C.
320 Kidney & Urinary Tract Infections Age GE 70 and/or C.C.
321 Kidney & Urinary Tract Infections Age 18-69 w/o C.C.
322 Kidney & Urinary Tract Infections Age 0-17
323 Urinary Stones Age GE 70 and/or C.C.
324 Urinary Stones Age LT 70 w/o C.C.
325 Kidney & Urinary Tract Signs & Symptoms Age GE 70 and/or C.C.
326 Kidney & Urinary Tract Signs & Symptoms Age 18-69 w/o C.C.
327 Kidney & Urinary Tract Signs & Symptoms Age 0-17
331 Other Kidney & Urinary Tract Diagnoses Age GE 70 and/or C.C.
332 Other Kidney & Urinary Tract Diagnoses Age 18-69 w/o C.C.
333 Other Kidney & Urinary Tract Diagnoses Age 0-17
338 Testes Procedures, for Malignancy
339 Testes Procedures, Non-malignant Age GE 18
340 Testes Procedures, Non-malignant Age 0-17
342 Circumcision Age GE 18
343 Circumcision Age 0-17
346 Malignancy, Male Reproductive System, Age GE 70 and/or C.C.
347 Malignancy, Male Reproductive System, Age LT 70 w/o C.C.
348 Benign Prostatic Hypertrophy Age GE 70 and/or C.C.
349 Benign Prostatic Hypertrophy Age LT 70 w/o C.C.
350 Inflammation of the Male Reproductive System
351 Sterilization, Male
352 Other Male Reproductive System Diagnoses
360 Vagina, Cervix & Vulva Procedures
361 Laparoscopy & Endoscopy (Female) Except Tubal Interruption
362 Laparoscopic Tubal Interruption
363 D&C, Conization & Radio-implant, for Malignancy
364 D&C, Conization Except for Malignancy
365 Other Female Reproductive System O.R. Procedures
368 Infections, Female Reproductive System
371 Cesarean Section w/o C.C.
372 Vaginal Delivery with Complicating Diagnoses
374 Vaginal Delivery with Sterilization and/or D&C
375 Vaginal Delivery with O.R. Proc Except Steril and/or D&C

376 Postpartum Diagnoses w/o O.R. Procedure
377 Postpartum Diagnoses with O.R. Procedure
378 Ectopic Pregnancy
379 Threatened Abortion
380 Abortion w/o D&C
381 Abortion with D&C
390 Neonates with Other Significant Problems
392 Splenectomy Age GE 18
393 Splenectomy Age 0-17
394 Other O.R. Procedures of the Blood & Blood Forming Organs
397 Coagulation Disorders
398 Reticuloendothelial & Immunity Disorders Age GE 70 and/or C.C.
399 Reticuloendothelial & Immunity Disorders Age LT 70 w/o C.C.
401 Lymphoma or Leukemia with Minor O.R. Proc Age GE 70 and/or C.C.
402 Lymphoma or Leukemia with Minor O.R. Procedure Age LT 70 w/o C.C.
403 Lymphoma or Leukemia Age GE 70 and/or C.C.
404 Lymphoma or Leukemia Age 18-69 w/o C.C.
405 Lymphoma or Leukemia Age 0-17
406 Myeloprolif Disord or Poorly Diff Neoplasm w Maj O.R. Proc & C.C.
407 Myeloprolif Disord or Poorly Diff Neopl w Maj O.R. Proc w/o C.C.
408 Myeloprolif Disord or Poorly Diff Neopl with Minor O.R. Proc
411 History of Malignancy w/o Endoscopy
412 History of Malignancy with Endoscopy
413 Other Myeloprolif Disord or Poorly Diff Neopl Dx Age GE 70 &/or C.C.
414 Other Myeloprolif Disord or Poorly Diff Neopl Dx Age LT 70 w/o C.C.
415 O.R. Procedure for Infectious & Parasitic Diseases
416 Septecemia Age GE 18
417 Septecemia Age 0-17
418 Postoperative & Post-Traumatic Infections
419 Fever of Unkown Origin Age GE 70 and/or C.C.
424 O.R. Procedures with Principal Diagnosis of Mental Illness
434 Drug Dependence
436 Alcohol Dependence
439 Skin Grafts for Injuries
440 Wound Debridements for Injuries
441 Hand Procedures for Injuries
443 Other O.R. Procedures for Injuries Age LT 70 w/o C.C.
449 Toxic Effect of Drugs Age GE 70 and/or C.C.
450 Toxic Effects of Drugs Age 18-69 w/o C.C.
451 Toxic Effects of Drugs Age 0-17
454 Other Injuries, Poisonings & Toxic Eff Diag Age GE 70 and/or C.C.
455 Other Injuries, Poisonings & Toxic Eff Diag Age LT 70 w/o C.C.
458 Non-extensive Burns with Skin Grafts
459 Non-extensive Burns with Wound Debridement & Other O.R. Proc
460 Non-extensive Burns w/o O.R. Procedure

001 Craniotomy Age GE 18 Except Trauma
002 Craniotomy for Trauma Age GE 18
003 Craniotomy Age LT 18
004 Spinal Procedures
005 Extracranial Vascular Procedures
009 Spinal Disorders & Injuries
010 Nervous System Neoplasms Age GE 70 and/or C.C.
011 Nervous System Neoplasms Age LT 70 and/or C.C.
014 Specific Cerebrovascular Disorders Except TIA
027 Traumatic Stupor & Coma, Coma GT 1 Hr
028 Traumatic Stupor & Coma, Coma LT 1 Hr Age GE 70 and/or C.C.
029 Traumatic Stupor & Coma LT 1 Hr Age 18-69 w/o C.C.
030 Traumatic Stupor & Coma LT 1 Hr Age 0-17
038 Primary Iris Procedures
039 Lens Procedures
040 Extraocular Procedures Except Orbit Age GE 18
041 Extraocular Procedures Except Orbit Age 0-17
042 Intraocular Procedures Except Retina, Iris & Lens
043 Hyphema
044 Acute Major Eye Infections
045 Neurological Eye Disorders
046 Other Disorders of the Eye Age GE 18 with C.C.
047 Other Disorders of the Eye Age GE 18 w/o C.C.
048 Other Disorders of the Eye Age 0-17
049 Major Head & Neck Procedures
064 Ear, Nose & Throat Malignancy
079 Respiratory Infections & Inflammations Age GE 70 and/or C.C.
080 Respiratory Infections & Inflammations Age 18-69 w/o C.C.
082 Respiratory Neoplasms
083 Major Chest Trauma Age GE 70 and/or C.C.
084 Major Chest Trauma Age LT 70 w/o C.C.
085 Pleural Effusion Age GE 70 and/or C.C.
086 Pleural Effusion Age LT 70 w/o C.C.
087 Pulmonary Edema & Respiratory Failure
088 Chronic Obstructive Pulmonary Disease
089 Simple Pneumonia & Pleurisy Age GE 70 and/or C.C.
092 Interstitial Lung Disease Age GE 70 and/or C.C.
094 Pneumothorax Age GE 70 and/or C.C.
095 Pneumothorax Age LT 70 w/o C.C.
114 Upper Limb & Toe Amputation for Circ System Disorder
115 Permanent Cardiac Pacemaker Implant with AMI or CHF
116 Permament Cardiac Pacemaker Implant w/o AMI or CHF
120 Other O.R. Procedures on the Circulatory System
121 Circulatory Disorders with AMI & C.V. Comp. Disch. Alive
122 Circulatory Disorders with AMI w/o C.V. Comp. Disch. Alive
126 Acute & Subacute Endocarditis
127 Heart Failure & Shock
130 Peripheral Vascular Disorders Age GE 70 and/or C.C.
131 Peripheral Vascular Disorders Age LT 70 w/o C.C.
132 Atherosclerosis Age GE 70 and/or C.C.
135 Cardiac Congenital & Valvular Disorders Age GE 70 and/or C.C.

136 Cardiac Congenital & Valvular Disorders Age 18-69 w/o C.C.
137 Cardiac Congenital & Valvular Disorders Age 0-17
143 Chest Pain
146 Rectal Resection Age GE 70 and/or C.C.
147 Rectal Resection Age LT 70 w/o C.C.
148 Major Small & Large Bowel Procedures Age GE 70 and/or C.C.
149 Major Samll & Large Bowel Procedures Age LT 70 w/o C.C.
150 Peritoneal Adhesiolysis Age GE 70 and/or C.C.
152 Minor Small & Large Bowel Procedures Age GE 70 and/or C.C.
154 Stomach, Esophageal & Duodenal Procedures Age GE 70 and/or C.C.
155 Stomach, Esophageal & Duodenal Procedures Age 18-69 w/o C.C.
156 Stomach, Esophageal & Duodenal Procedures Age 0-17
157 Anal Procedures Age GE 70 and/or C.C.
164 Appendectomy with Complicated Princ. Diag Age GE 70 and/or C.C.
172 Digestive Malignancy Age GE 70 and/or C.C.
173 Digestive Malignancy Age LT 70 w/o C.C.
175 G.I. Hemorrhage Age LT 70 w/o C.C.
176 Complicated Peptic Ulcer
179 Inflammatory Bowel Disease
180 G.I. Obstruction Age GE 70 and/or C.C.
181 G.I. Obstruction Age LT 70 w/o C.C.
193 Biliary Tract Proc Exc Tot Cholecystectomy Age GE 70 &/or C.C.
194 Biliary tract Proc Exc Tot Cholecystectomy Age LT 70 w/o C.C.
195 Total Cholecystectomy with C.D.E. Age GE 70 and/or C.C.
196 Total Cholecystectomy with C.D.E. Age LT 70 w/o C.C.
197 Total Cholecystectomy w/o C.D.E. Age GE 70 and/or C.C.
198 Total Cholecystectomy w/o C.D.E. Age LT 70 w/o C.C.
199 Hepatobiliary Diagnostic Procedure for Malignancy
200 Hepatobiliary Diagnostic Procedure for Non-Malignancy
201 Other Hepatobiliary or Pancreas O.R. Procedures
202 Cirrhosis & Alcoholic Hepatitis
204 Disorders of Pancreas Except Malignancy
205 Disorders of Liver Exc Malig.Cirr.Alc Hepa Age GE 70 and/or C.C.
206 Disorders of Liver Exc Malig.Cirr.Alc Hepa Age LT 70 w/o C.C.
207 Disorders of the Biliary Tract Age GE 70 and/or C.C.
208 Disorders of the Biliary Tract Age LT 70 w/o C.C.
212 Hip & Femur Procedures Except Major Joint Age 0-17
213 Amputations for Musculoskeletal System & Conn. Tissue Disorders
214 Back & Neck Procedures Age GE 70 and/or C.C.
215 Back & Neck Procedures Age LT 70 w/o C.C.
218 Lower Extrem & Humer Proc Exc Hip,Foot,Femur Age GE 70 &/or C.C.
219 Lower Extrem & Humer Proc Exc Hip,Foot,Femur Age 18-69 w/o C.C.
220 Lower Extrem & Humer Proc Exc Hip,Foot,Femur Age 0-17
221 Knee Procedures Age GE 70 and/or C.C.
222 Knee Procedures Age LT 70 w/o C.C.
223 Upper Extremity Proc Exc Humerus & Hand Age GE 70 and/or C.C.
224 Upper Extremity Proc Exc Humerus & Hand Age LT 70 w/o C.C.
257 Total Mastectomy for Malignancy Age GE 70 and/or C.C.
258 Total Mastectomy for Malignancy Age LT 70 w/o C.C.
259 Subtotal Mastectomy for Malignancy Age GE 70 and or C.C.
260 Subtotal Mastectomy for Malignancy Age LT 70

274 Malignant Breast Disorders Age GE 70 and/or C.C.
275 Malignant Breast Disorders Age LT 70 w/o C.C.
285 Amputations for Endocrine, Nutritional & Metabolic Disorders
286 Adrenal & Pituitary Procedures
290 Thyroid Procedures
305 Kidney, Ureter & Maj Bldr Proc for Non-malig Age LT 70 w/o C.C.
306 Prostatectomy Age GE 70 and/or C.C.
307 Prostatectomy Age LT 70 w/o C.C.
310 Transurethral Procedures Age GE 70 and/or C.C.
315 Other Kidney & Urinary Tract O.R. Procedures
328 Urethral Stricture Age GE 70 and/or C.C.
329 Urethral Stricture Age 18-69 w/o C.C.
330 Urethral Stricture Age 0-17
334 Major Male Pelvic Procedures with C.C.
335 Major Male Pelvic Procedures w/o C.C.
336 Transurethral Prostatectomy Age GE 70 and/or C.C.
337 Tranurethral Prostatectomy Age LT 70 w/o C.C.
341 Penis Procedure
344 Other Male Reproductive System O.R. Procedures for Malignancy
345 Other Male Reproductive System O.R. Procedures Except for Malig
354 Non-radical Hysterectomy Age GE 70 and/or C.C.
355 Non-radical Hysterectomy Age LT 70 w/o C.C.
356 Female Reproductive System Reconstructive Procedures
357 Uterus & Adenexa Procedures, for Malignancy
358 Uterus & Adenexa Proc for Non-malignancy Except Tubal Interrupt
359 Tubal Interruption for Non-malignancy
366 Malignancy, Female Reproductive System Age GE 70 and/or C.C.
367 Malignancy, Female Reproductive System Age LT 70 w/o C.C.
370 Cesarean Section with C.C.
373 Vaginal Delivery w/o Complicating Diagnoses
388 Prematurity w/o Major Problems
389 Full Term Neonate with Major Problems
400 Lymphoma or Leukemia with Major O.R. Procedure
409 Radiotherapy
410 Chemotherapy
442 Other O.R. Procedures for Injuries Age GE 70 and/or C.C.
444 Multiple Trauma Age GE 70 and/or C.C.
445 Multiple Trauma Age 18-69 w/o C.C.
446 Multiple Trauma Age 0-17
453 Complications of Treatment Age LT 70 w/o C.C.

DRG LEVEL 4

075 Major Chest Procedures
076 O.R. Proc on the Resp System Except Major Chest with C.C.
077 O.R. Proc on the Resp System Except Major Chest w/o C.C.
078 Pulmonary Embolism
103 Heart Transplant
104 Cardiac Valve Procedure with Pump & with Cardiac Cath
105 Cardiac Valve Procedure with Pump & w/o Cardiac Cath

106 Coronary Bypass with Cardiac Cath
107 Coronary Bypass w/o Cardiac Cath
108 Cardiothor Proc, Except Valve & Coronary Bypass with Pump
109 Cardiothoracic Procedures w/o Pump
110 Major Reconstructive Vascular Procedures Age GE 70 and/or C.C.
111 Major Reconstructive Vascular Procedures Age LT 70 w/o C.C.
112 Vascular Procedures Except Major Reconstruction
113 Amputation for Circ System Disorders Except Upper Limb & Toe
123 Circulatory Disorders with AMI, Expired
124 Circulatory Disorders Exc AMI, with Card Cath & Complex Diag
125 Circulatory Disorders Exc AMI, with Card Cath w/o Complex Diag
129 Cardiac Arrest
138 Cardiac Arrhythmia & Conduction Disorders Age GE 70 and/or C.C.
139 Cardiac Arrhythmia & Conduction Disorders Age LT 70 w/o C.C.
141 Syncope & Collapse Age GE 70 and/or C.C.
142 Syncope & Collapse Age LT 70 w/o C.C.
174 G.I. Hemorrhage Age GE 70 and/or C.C.
191 Major Pancreas, Liver & Shunt Procedures
192 Minor Pancreas, Liver & Shunt Procedures
203 Malignancy of Hepatobiliary or Pancreas O.R. Procedures
209 Major Joint Procedures
210 Hip & Femur Procedures Except Major Joint Age GE 70 and/or C.C.
211 Hip & Femur Procedures Except Major Joint Age 18-69 w/o C.C.
302 Kidney Transplant
303 Kidney, Ureter & Major Bladder Proceudre for Malignancy
304 Kidney, Ureter & Maj Bldr Proc for Non-Malig Age GE 70 &/or C.C.
353 Pelvic Evisceration, Radical Hysterectomy & Vulvectomy
385 Neonates, Died or Transferred
386 Extreme Immaturity, Neonate
387 Prematurity with Major Problems
452 Complications of Treatment Age GE 70 and/or C.C.
456 Burns, Transferred to Another Acute Care Facility
457 Extensive Burns

Standardized rates and computing in SAS RTI

This appendix explains how we computed the standardized rate (AE, for example) for a subgroup of the population, such as blacks; when there were several subgroups, we let k index the subgroups. We are standardizing with respect to some other variable or cross classification of other variables (i.e., some other covariates), where the cross-classification has M levels, m=1,...,M. We will refer to the m^{th} level as a strata.

Then, let

$\hat{R}_{km}$ = ratio estimate of the rate for the m^{th} strata within the k^{th} subgroup

$\hat{Var}(\hat{R}_{km})$ = estimated variance of $\hat{R}_{km}$ (from Proc Sesudaan);

w_m = weight given to the m^{th} strata for standardization, $\sum_{m=1}^{M} w_m = 1$. For our data, $w_m = N_m/N$, where

N_m = the number of population elements in stratum m;

$\sum_{m=1}^{M} N_m = N$ = the number of elements in the population.

Then, the estimated standardized rate for the k^{th} group is

$$\hat{R}_k^s = \sum_{m=1}^{M} w_k \hat{R}_{km} ,$$

and

$$\hat{Var}(\hat{R}_k^s) = \sum_{m=1}^{M} w_m^2 \hat{Var}(\hat{R}_{km}) + 2 \sum_{q<m} \sum w_q w_k \hat{Cov}(\hat{R}_{kq}, \hat{R}_{km}) \qquad (2)$$

Because of the cluster design, $\hat{R}_{kq}$ and $\hat{R}_{km}$ may be correlated since individuals from the same hospital (i.e., cluster) may contribute to both $\hat{R}_{kq}$ and $\hat{R}_{km}$.

I. Calculating $\hat{Var}(\hat{R}^s_k)$

To calculate $\hat{Var}(\hat{R}^s_k)$, Proc Sesudaan first calculates the variable

$$z_{hij}(km) = \{W_{hij}[Y_{hij} - \hat{R}_{km}]/x_{km}\}km_{hij} \ , \qquad (1)$$

as in appendix 5.V.3, with km the indexing covariate factor level. Proc Sesudaan then calculates the modified variable

$$z^s_{hij}(k) = \sum_{m=1}^{M} w_m z_{hij}(km) \ ;$$

to estimate the variance, Proc Sesudaan plugs $z^s_{hij}(k)$ into equation (6) of appendix 5.V.3. Because of the cluster design $\hat{R}^s_{k_1}$ and $\hat{R}^s_{k_2}$, $k_1 \neq k_2$, may be correlated.

$\hat{Cov}(\hat{R}^s_{k_1}, \hat{R}^s_{k_2})$ is calculated indirectly from Proc Sesudaan as in equation (15) in appendix 5.V.3.

II. Wald test

We want to perform a Wald test for

H_o: $R_1^s = R_2^s = \cdots = R_D^s$ (i.e., all the standardized rates are equal). We can also write this null hypothesis as $C\underline{R}^s = \underline{0}$, for the appropriate contrast matrix C, and where $\underline{R}^s = [R_1^s, \ldots, R_D^s]'$. We can use the $\hat{\underline{R}}^s = \{\hat{R}_k^s\}$ to test this null

hypothesis. Again,

because of the clustering $\hat{R}^s_{k_1}$ and $\hat{R}^s_{k_2}$, $k_1 \neq k_2$, are correlated. We can estimate the covariance matrix with $\hat{Var}(\hat{\underline{R}}^s)$ obtained by the method outlined in part I of this appendix, and use as our test statistic

$$G^2 = (C\hat{\underline{R}}^s)'[C\hat{Var}(\hat{\underline{R}}^s)C'](C\hat{\underline{R}}^s) ,$$

which, under the null hypothesis, is approximately chi-square with (D-1) degrees of freedom.

III. Compute standardized AE and negligent AE rates

1) race

We computed AE and negligent AE standardized rates for the three different levels of the variable race:

i) blacks
ii) hispanics
iii) whites/others.

We chose to standardize on 17 age by drg (see Technical Appendix 5.V.4 for a discussion of drg level) strata:

a. Initially, we formed 20 standardization strata by the cross-classification of age (5 levels: newborn,0-15,16-44,45-64,65$^+$) and DRG (4 levels). Then, we formed the 17 standardization strata by collapsing those 65+ into one strata. The standardization weights are given in table I.

2) hospital characteristics

We computed AE and negligent AE standardized rates for different variables which characterize hospitals. We looked at each of the following variables separately (in this sense, each variable is analyzed as a separate covariate, just like race). The same 17 age by drg standardization weights were used for five of these variables.

A. Hospital race mix, with three levels

i) > 83% minority
ii) 15 - 83% minority
iii) < 15% minority

B. Hospital size, with three levels

i) 0-7999 discharges
ii) 8000-19,999 discharges
iii) 20,000+ discharges

C. Hospital teaching status, with three levels

i) non-teaching
ii) teaching
iii) primary teaching

D. Location of Hospital with four levels

i) Upstate, non-MSA
ii) Nassau-Suffolk-Westchester
iii) New York City
iv) Upstate MSA

3) Hospital ownership

We computed AE and negligent AE standardized rates for the three different levels of the variable hospital ownership:

i) non-profit
ii) profit
iii) Government and HHC

Profit hospitals cared for few people under age 15, so two sets of analyses were done. The first excluded the profit hospitals and compared non-profit and government hospitals, standardizing by the 17 age and DRG strata above. In the second set of analyses, those under 15 were excluded from the analyses. From this, we chose 9 standardization strata, formed by initially creating 12 strata (the cross-classification of age (3 levels: 16-44, 45-64, 65^{+}) and DRG (4 levels)) and then collapsing those over 65 into one stratum. These standardization weights are also given in table I.

4) reimbursement

We computed AE and negligent AE standardized rates for the four different levels of the variable reimbursement:

i) self-pay
ii) medicare
iii) medicaid
iv) BC/Others .

Medicare was mostly restricted to individuals over 45, so we first compared the 3 levels i), iii), and iv) using the 17 standardization strata discussed in 1.a. above (however, we recalculated the standardization weights as the population age by drg proportions for non-medicaid individuals). These weights are also given in table I.

Then, we restricted our attention to individuals 45 and older, in which we compare all 4 levels. We chose 5 standardization strata, formed by initially creating 8 strata (the cross-classification of age (2 levels: 45-64,65+) and DRG (4 levels)) and then collapsing those over 65 into one stratum. These standardization weights are also given in table I.

5) Individual Hospitals

We computed standardized AE rates for the 19 hospitals with 900 or more discharges. Because of small numbers in the younger age groups, we combined the newborn and under 15 age groups. We chose 13 standardization strata, formed by initially creating 16 strata (the cross-classification of age (4 levels: 0-15, 16-44, 45-64, 65^{+}) and DRG (4 levels)) and then collapsing those over 65 into one stratum. These standardization weights are also given in table I.

6) age

We computed AE and negligent AE standardized rates for the five different levels of the variable age:

i) newborn
ii) 0-15
iii) 16-44
iv) 45-64
v) 65+

We standardized age by drg (4 drg levels). The standardization weights are given in table II.

Table I. Weights (in %) for standardizing by age and drg

age	drg	All population (for race & hospital char-acteristics)	All population over 55 (for hospital ownership)	Non-medicare population (for reimbursement)
newborn	1	6.26	0.00	8.78
	2	1.66	0.00	2.32
	3	1.13	0.00	1.59
	4	0.44	0.00	0.62
0-15	1	1.28	0.00	1.79
	2	5.93	0.00	8.30
	3	1.08	0.00	1.51
	4	0.24	0.00	0.33
16-44	1	3.30	6.10	4.57
	2	18.64	25.40	25.79
	3	11.69	14.20	16.20
	4	0.71	1.50	0.93
45-64	1	1.19	1.50	1.45
	2	9.64	14.40	11.94
	3	7.51	11.00	9.36
	4	2.09	4.50	2.64
65+	all	27.21	21.20	1.88

Table I. Weights (in %) for standardizing by age and drg (cont'd)

age	drg	All population (for reimbursement)	All population (for individual hospitals)
newborn	1	0.00	0.00
	2	0.00	0.00
	3	0.00	0.00
	4	0.00	0.00
0-15	1	0.00	4.10
	2	0.00	11.30
	3	0.00	5.00
	4	0.00	1.70
16-44	1	0.00	4.70
	2	0.00	19.80
	3	0.00	11.10
	4	0.00	1.20
45-64	1	2.51	1.20
	2	20.24	11.10
	3	15.75	8.60
	4	4.39	3.50
65^{+}	all	57.11	16.50

Table II. Weights for standardizing by drg

drg	All population (for age)
1	13.98
2	45.60
3	33.71
4	6.71

IV. Computing standardizing rates of negligent AE given AE

We found that the rates of negligent AE given AE do not depend on drg (see Chapter 5, Table 5.6). Thus, in computing standardized rates for negligent AE given AE, we did not standardize on drg; the rates of negligent AE given AE were computed as follows:

i) for race, hospital characteristics, and hospital ownership we used two age standardization strata (less than 65, 65 and over), as given in table III;

ii) for reimbursement, we compared self-pay, medicaid, and bc/others using two age standardization strata (less than 65, 65 and over), with weights calculated as the population age proportions for non-medicare individuals (see table IV);

iii) for reimbursement, we compared all four groups for individuals over 44 using two age standardization strata (45-64, and 65+), again, see table IV;

iv) for age, we computed unstandardized (crude) rates .

Table III. Weights for standardizing negligent AE given AE rates by age

age	All population (for race, hospital characteristics & ownership)	Non-medicare population (for reimbursement)	All population over 44 (for reimbursement)
< 65	72.79	98.12	42.89
65+	27.21	1.88	57.11

COMPARING FOLLOW-UP AND REPORTED AE RATES

This appendix outlines the methods used to compare the rate of adverse events (AEs) at follow-up with the rate obtained in the original record review. The follow-up survey was examined in 3 ways:

1. The difference between original and follow-up rates for each of the six hospitals was computed.
2. The average follow-up rate for the six hospitals was compared with the original average rate for those hospitals.
3. The estimated rate for all missing records was compared with the estimated population rate for the original review.

1. For comparison of original and follow-up AE rates in individual hospitals, the weighted difference was computed. The weight for the original survey was equal to the inverse of the probability of a record being sampled in that hospital. (See Chapter 4 for additional information.)

$$wt_i = SU_j * STR_m * C$$

where wt_i = the weight for the i^{th} record
SU_j = the sampling units for the j^{th} hospital
STR_m = the over/under sampling factor for the m^{th} stratum
c = adjustment constant (= 1.1)

The weight for follow-up records was adjusted to compensate for records not found. The effect was to apply the AE rate in the records found to the records still missing.

$$wt_{Fi} = wt_i * \frac{\sum_{i \epsilon M_h} wt_i}{\sum_{i \epsilon F_h} wt_i}$$

where wt_{Fi} = weight for the i^{th} found record

M_h = the set of all records i which were missing on first review in hospital h

F_h = the set of all records i which were found on second review in hospital h

so the difference between the follow-up rate and the original rate for each hospital is

$$dh = \frac{\sum_{i \epsilon F_h} wt_{Fi} * ae_i}{\sum_{i \epsilon F_h} wt_{Fi}} - \frac{\sum_{i \epsilon G_h} wt_i * ae_i}{\sum_{i \epsilon G_h} wt_i}$$

where ae_i = 1 if record i is an AE

= 0 if record i is not an AE

G_h = the set of all records i which were examined in the first review in hospital h

2. The difference between the average follow-up AE rate and the original AE rate for the six hospitals was computed using the same weights as in (1).

$$D = \frac{\sum_{h=1}^{6} \sum_{i \epsilon F_h} wt_{Fi} * ae_i}{\sum_{h=1}^{6} \sum_{i \epsilon F_h} wt_{Fi}} - \frac{\sum_{h=1}^{6} \sum_{i \epsilon G_h} wt_i * ae_i}{\sum_{h=1}^{6} \sum_{i \epsilon G_h} wt_i}$$

3. For comparison of the estimated rate of all missing records to the estimated population rate based on original records, the method of computing ratio estimates and standard errors described in Appendix 5.V.3 was used.

For records found in the original review, the weight used was the same as used for analyses of patient and hospital characteristics (Chapter 4), excluding the adjustment for missing records.

$$V_i = \frac{86}{SU} * SU * STR_m * C * W_i$$

where w_i = adjustment for multiple methods of selection of a hospital because of the presence of undersized facilities.

V_i = the weight used

For records found on follow-up, this weight was adjusted for the records that remained missing and for the probability of the hospital being selected for follow-up review.

$$V_{Fi} = v_i * \frac{\sum_{i \epsilon M_h} v_i}{\sum_{i \epsilon F_h} v_i} * \frac{\sum_{i=1}^{6} \sum_{i \epsilon M_h} v_i}{\sum_{i \epsilon M_n} v_i} * \frac{1}{P}$$

where P = the probability of a hospital being selected

DETERMINING DISABILITY SCORES FROM MULTIPLE REVIEWS

I. Introduction

This technical appendix describes our methods for determining both population estimates of patients' disabilities and case-specific disability scores.

As Chapter 5 explains, the medical record review protocol required each physician-reviewer who found an adverse event (AE) to assess the degree of the patient's disability caused by the adverse event over and above the patient's disability from underlying disease. An adverse event by definition must include either a prolongation of hospitalization or a disability. Possible scores ranged from 1 (minimal impairment) to 8 (death), with a score of 9 indicating that the reviewer could not reasonably judge disability from the medical record.

Because virtually all cases received two independent reviews, as dictated by the study design, the two reviewing physicians could and did disagree on the extent of disability. In 708 of the 1,085 (65.3%) adverse events for which there were two reviews, the physicians agreed exactly on the disability score. Disagreement among reviewers included instances of two different scores in the range of 1 to 8 as well as one determinate score in this range and the other indeterminate (score = 9).

For a series of comparisons of adverse events by patient disability, handling two different disability scores became unwieldy. The categories are not easily ordered so that averaging would produce meaningless results. Also, where one reviewer could assess patient disability and the other could not, we felt that a single review would be sufficient to assign a score to the case. For statewide estimates of the numbers of adverse events by disability, we chose to preserve the indeterminate reviews and report the percentage of cases with scores equal to 9.

To serve the different estimation needs of the study we therefore devised two systems for determining the distribution of cases by patient disability. The two systems might assign different scores to a single case only for the one third of the adverse events in which reviewers disagreed.

II. Methods

A. Disability By Review

For purposes of determining state wide estimates of adverse event rates by disability, we analyzed our data by review. Of the 1,133 adverse events, we had 1,085 cases with two reviews and 48 with a single review, giving a total of 2,218 reviews. For the cases with two reviews, we assigned one half of the case weight to each review; for those with one review we assigned the full case weight to the review. The sum of the weights for the 2218 reviews then equalled the sum of the weights of the cases (2,671,863). Using this method we calculated ratio estimators for adverse event rates by disability subclass, as if there were 2218 adverse events, by the methods outlined in Technical Appendix 5.V.3. We were thus able to report both point estimators and variances for adverse event rates and population totals by disability (see Chapter 6, Table 6.5). This method of disability estimation also applied to estimates in Chapter 7 of state wide totals by patient disability (see Table 7.12 and 7.13).

B. Disability By Case

For other comparisons in Chapter 6, we found a need to assign to each case a specific disability score. We adopted the following algorithm for choosing a single disability score in the one third of cases in which reviewers disagreed:

1. Where one reviewer either could not determine disability (score = 9) or neglected to answer the question (2 cases), the case received the score of the other reviewer.

2. Where both reviewers could not determine disability (score = 9), the case received a score of 1 (minimal impairment).

3. In the one case with 2 reviews, one indeterminate (score = 9) and the other missing, we imputed a value for disability as described in Technical Appendix 5.V.1. This practice prevented the possibility of indeterminate disability scores for any cases.

4. In all other cases in which the two reviewers disagreed on disability, we chose one reviewer's assessment at random and assigned it to the case.

III. Results

These two methods of computing disability produced the following distributions (weighted) of disability scores;

Method of Review

Disability	By Review (Method A)	By Case (Method B)
1	56.8	59.9
2	11.0	12.7
3	2.7	2.8
4	2.8	2.9
5	3.9	4.2
6	1.3	1.2
7	1.3	1.2
8	13.6	14.7
9	6.6	0.0

The differences in percentages result largely from the absence of undetermined disability in the analysis by case.

Chapter 6

THE NATURE AND INCIDENCE OF MEDICAL INJURY: RESULTS OF THE MEDICAL RECORD REVIEW

Summary of Findings of the Record Review

1. Incidence of Adverse Events

Our analysis is based on 30,121 of the 31,429 records selected for the study sample (96%). After preliminary screening, physicians reviewed 7,743 records, from which a total of 1,133 adverse events (AEs) were identified that occurred as a result of medical management in the hospital or required hospitalization for treatment. Of this group, 280 were judged to result from negligent care. Weighting these figures according to the sample plan, we estimated the incidence of AEs for hospitalizations in New York in 1984 to be 3.7%. Of these AEs, 27.6% were due to negligence, or 1.0% of all hospital discharges had an AE. The rate of AEs is slightly lower than that reported in the California Medical Association study of 1975, but our percentage due to negligence is higher, giving an overall incidence rate for negligent AEs similar to the California Study.

Physician confidence in their judgments of causation of AEs spanned a broad range, but only 1.3% of all discharges were in the close-call range (AE confidence score 3 or 4). An even smaller fraction, 0.7% of all discharges were close call negligent events but they contributed a larger proportion of total negligent events.

The majority of AEs (57%) resulted in minimal and transient disability, but 14% of patients died at least in part as a result of their AE, and in another 9% the disability lasted longer than 6 months. Based on these figures, we estimate that 13,451 deaths and 2,550 cases of permanent total disability resulted from medical injury in New York hospitals in 1984. Negligent AEs

resulted in greater disability and were associated with 51% of the deaths from medical injury.

2. Risk Factors -- Individual Characteristics

The risk of sustaining an AE increased with age. When rates were standardized for DRG level, persons over 65 had twice the chance of sustaining an AE than those 16-44 years. Newborns had half the AE rate of age 16-44 group. The percentage of AEs resulting from negligence was increased only in patients over 65. No significant gender differences occurred in AE or negligence rates. We found no significant differences in AE rates by payer status. The rate of negligence was high among the self-pay group but only marginally statistically significant and the higher rate disappears when payer groups are compared among whites only. Although blacks had a higher standardized rate of AEs over all and AEs resulting from negligence, racial differences were not statistically significant. The higher rates among blacks appeared to be due to care provided at hospitals with a higher proportion of minority patients. Both AE rates and percentage of AEs due to negligence were significantly increased in patients treated at these hospitals. At hospitals that care for a mix of races, blacks and whites had nearly identical AE rates and percentage of AEs due to negligence.

3. Risk Factors -- Hospital Characteristics

AE rates varied tenfold between individual hospitals when standardized for age and DRG level. Although standardized AE and negligence rates for small hospitals (<8,000 discharges/year) were less than for larger hospitals, these differences are only marginally significant. Hospital ownership (private, nonprofit, or government) also was not associated with significantly different rates of AEs. The fraction of AEs due to negligence in

government hospitals was 50% higher than in nonprofit institutions, however, and three times that in proprietary hospitals. These differences are significant. The standardized rate of AEs in upstate non-MSA hospitals was one third that of upstate metropolitan hospitals and less than one fourth that in New York City. These differences are highly significant. The percentage of AEs due to negligence was not significantly different. Non-teaching hospitals had half the AE rates of university or affiliated teaching hospitals. On the other hand, university teaching hospitals had rates of negligence that were less than half those of the non-teaching or affiliated hospitals.

4. Nature of Adverse Events

Nearly half (47%) of AEs occurred in patients undergoing surgery, but the percentage of AEs caused by negligence was higher in nonsurgical AEs (37% versus 17%). AEs resulting from an error in diagnosis and from non-invasive treatment were judged to be due to negligence in over three fourths of patients. The high rate of AEs in patients over 65 was related to three types of AEs: nontechnical postoperative complications, complications of noninvasive therapy, and falls. A larger proportion of AEs in younger patients were due to surgical failures. Not surprisingly, the operating room was the site of management for the highest fraction of AEs but had a low negligence rate. The proportion of negligent AEs was highest in the emergency room (70% of all AEs in the ER).

The most common error resulting in an AE was an error of performance of a procedure or operation, but diagnostic errors and prevention errors were more likely to be judged negligent and to result in serious disability.

Increasing severity (gravity) of negligence was associated with a greater proportion of AEs causing serious disability (moderate impairment with recovery taking more than six months, permanent disability, or death). Nearly three fourths of patients with an AE judged gravely negligent suffered a serious disability greater than level 2, more than three times the proportion for non-negligent AEs. Mortality in this group was over six times that for non-negligent AEs.

While many of the AEs we found, such as drug reactions, are not preventable with current knowledge, many others are. Indeed, a major reason for the dramatic effectiveness and safety of many current-day medical treatments is past success in reducing or eliminating complications analogous to those we have identified as AEs in this study: the high rates of heart block, bleeding, and mortality after heart surgery in the early years; the technical and immunologic problems associated with the early attempts at organ transplantation; the side effects of many drugs, and so on. These were the AEs of former decades. Their elimination resulted from increased understanding of their causes and from effective dissemination of the new knowledge.

Our listing of the types of AEs can be thought of, therefore, as an agenda for quality improvement. Some of the AEs we have identified could be prevented by better dissemination of information and establishment of protocols and guidelines for care. Our evidence suggests that this would be especially so for care given in the emergency room. Other AEs represent an educational, and in some cases, psychological challenge: how to effectively alter physician behavior. Still others may require improvements in the systems of care.

I. Introduction

This chapter presents the results of record review for a sample of patients hospitalized in New York in 1984. The objectives of the review were:

1. To measure the incidence of adverse events (AEs) and AEs due to negligence in hospitalized patients;
2. To identify a group of patients who had suffered AEs for purposes of determining the costs of the injuries;
3. To obtain information on the nature of AEs:
   a. the types of injury,
   b. the characteristics of the victims, and
   c. the factors that affect the risk of sustaining injury.

The rationale for the review, theoretical concerns in its execution, and methodological issues associated with clinical judgments of injury and negligence are presented in the preceding chapters and their appendices. Here we first present our estimates of the incidence of AEs and of negligent AEs. Next, we provide information on the extent of disability from medical injury and its relationship to negligent care as estimated from the medical records. (Additional data on disability, provided by patients, is presented in Chapter 8.)

We follow with an analysis of several risk factors for AEs: age, gender, race, payer status, and racial mix of hospital population. We then examine the role of hospital factors (size, ownership, location, and teaching status) in the occurrence of AEs and negligence. Next we offer a classification of AEs and explore the distribution of the various types of AEs and negligent AEs over all, by age groups, and by location. Then we look at the types of errors associated with AEs. Finally, we examine the association between gravity of negligence and severity of disability.

II. Overall Incidence

A. Population Estimates

A primary goal of our research was to develop an estimate of the incidence of AEs and of those AEs that were the result of negligent care. To accomplish this task, we identified all AEs in our sample. Because an AE may generate more than one hospitalization, we labeled the sampled discharge as the index hospitalization. We were interested in all AEs that occurred or were first discovered during the index hospitalization. Moreover, we counted as AEs those caused by medical management that occurred during the index hospitalization but discovered during a subsequent hospitalization or in subsequent outpatient care at the index hospital.

As discussed in Chapter 3, we restricted our sample to hospital records. We completed the screening process on 30,195 of the 31,429 records in the original random sample, or 96.1% of the records we requested. Of these, 7,817 met positive screening criteria, and we reviewed 7,743 of these at the second level review. Our analysis is based on 30,121 cases with negative screen or completed physician review. As explained in the text of Chapter 5 and in Technical Appendix 5.V.6, we adjusted the sample weights to account for the 1,234 missing records and the 74 records missing physician review.

Physician review uncovered 1,278 AEs and 306 negligent AEs (Figure 6.1).

Figure 6.1
Record Review Process

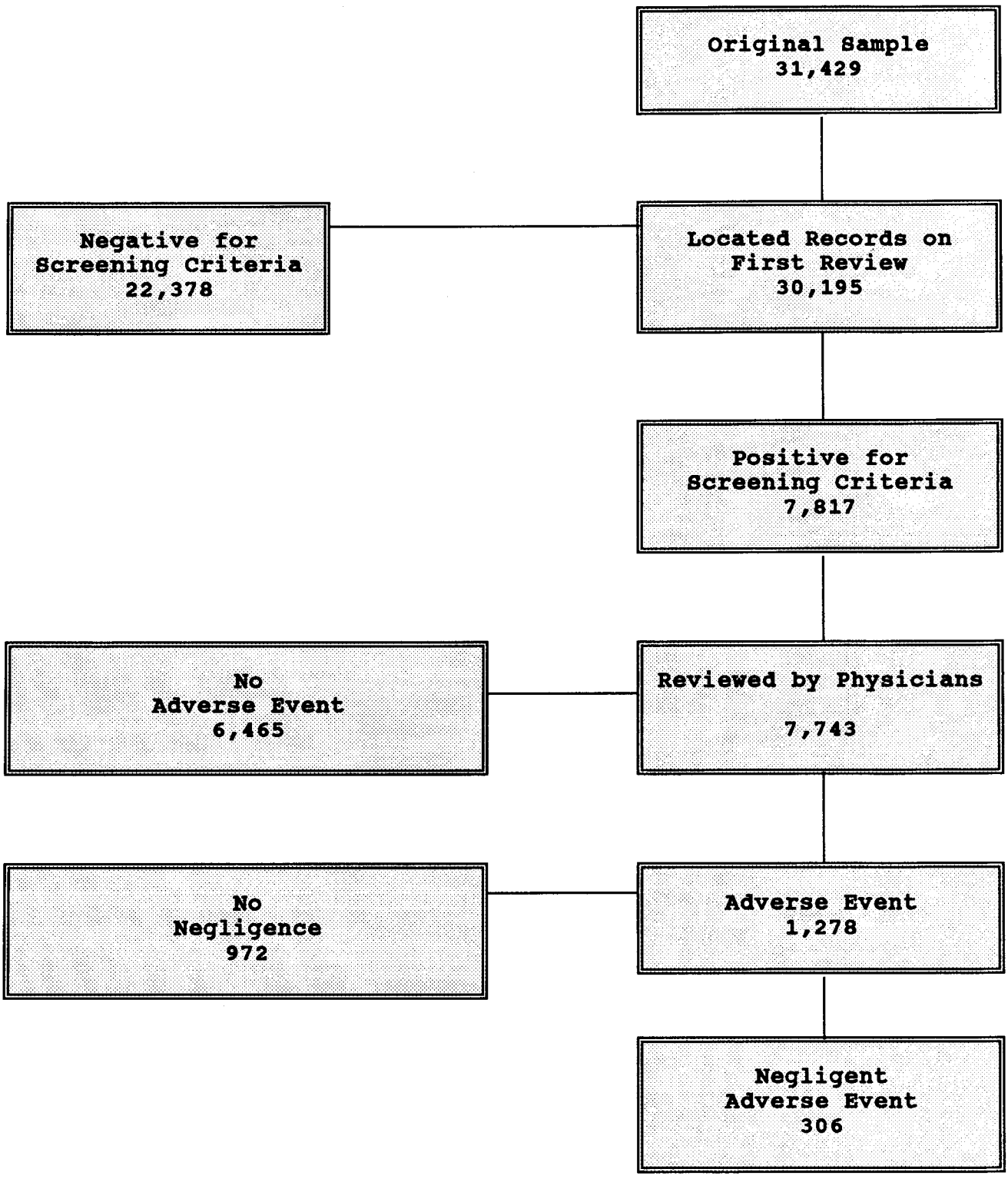


Table 6.1 presents the timing of the medical management and discovery of AEs and AEs due to negligence in each category, incorporating both raw counts and the population estimates.

Table 6.1

Incidence Categories for Adverse Events and Negligence

Category	Timing	Adverse Events Number	Adverse Events %	Adverse Events Due to Negligence Number	Adverse Events Due to Negligence %
1	AE from Medical Management in Index Hospitalization; discovered during Index Hospitalization	647[1] 55,046[2]	50.6 49.4	156 15,257	51.0 51.2
2	AE from Medical Management in Index Hospitalization; discovered in subsequent Outpatient management	78 6,327	6.1 5.7	7 776	2.3 2.6
3	AE from Medical Management in Index Hospitalization; discovered in subsequent Hospitalization	67 6,526	5.2 5.9	19 1,857	6.2 6.2
4	AE from Medical Management as Outpatient prior to Index Hospitalization; First Discovered during Index Hospitalization	167 16,142	13.1 14.5	59 6,019	19.3 20.2
5	AE from Medical Management in Hospitalization prior to Index Hospitalization; First Discovered during Index Hospitalization	319 27,420	25.0 24.6	65 5,903	21.2 19.8

[1] Sample count
[2] Weighted population total

For purposes of estimating the annual incidence of AEs, we cannot count all the AEs we identified because we would then overcount the number of AEs (as discussed in Chapter 5, Section III). Therefore, we restricted our definition of incidence to those that occurred and were discovered during the index hospitalization (category 1); those that resulted from prior outpatient medical management, but were first discovered during

the index hospitalization (category 4), and those that resulted from medical management in a prior hospitalization but were first discovered during the index hospitalization (category 5). (When we compute rates of AEs and negligence among hospitals, we use categories 1, 2 and 3 because those are the AEs we can associate with a particular hospital.)

There were 647 AEs that occurred during the index hospitalization and were first discovered in the records from that hospitalization (category 1). Another 486 occurred prior to the index hospitalization either in an outpatient setting (category 4) or at another hospital and were first discovered during the index hospitalization (category 5). This means that for purposes of calculating incidence, there were 1,133 AEs. Of these 1,133 AEs, 280 were judged to have arisen from negligent care.

These counts were used to develop the population estimates of rates of AEs and negligent AEs in New York in 1984. In order to estimate the number of AEs in all patients discharged from hospitals in New York, we must multiply each event observed in our sample by a weight proportional to the inverse of its probability of being sampled and then sum the products (Technical Appendix 5.V.6). As discussed in Chapter 4, we did not use a simple random sample of records within hospitals. Instead, we used a stratified sample with different sampling probabilities in different strata in order to increase the number of AEs sampled. As a result, not all discharges have the same weight.

Using the sample data, we estimate that among the 2,671,863 patients discharged from New York hospitals in 1984, 98,609 AEs and 27,179 negligent AEs occurred. With these figures, we estimated from our sample that the statewide incidence rate of AEs was 3.7% and the negligent AE rate was 1.0%; the percentage of AEs due to negligence was 27.6%.

The 95% confidence intervals (the results of 19 out of 20 similar studies would lie within this range) are given below.

	Estimate	95% Confidence Interval
Adverse Events	3.7%	3.2 to 4.2
Negligent AEs	1.0	0.8 to 1.2
% Negligent AEs	27.6	22.5 to 32.6

These estimates were based on a stratified random sample of all discharges, so they are likely to be more representative of true population rates than results from earlier studies by others that were restricted to selected hospitals. Although any sampling process has limitations, our evaluations of the sensitivity of the population estimates to the study design and implementation suggest that the population estimates are accurate. In the following sections we discuss these influences.

Before doing so, however, we compare our estimates with those of the California Medical Association (CMA) study completed ten years ago. In the sample of 20,864 records, the CMA investigators found 970 potentially compensable events PCEs for an overall rate of 4.65%. This rate is 26% greater than our sample estimate of 3.7%. In California, only 0.79% of records had evidence of negligence. This is 21% less than the results of our record review.

Several explanations for these differences may be put forward. One is that the concept of a potentially compensable event is different from that of an AE, and that the California investigators' use of the term negligence is different from ours. Using this explanation the differences in rates would stem from the measurement of different outcomes by the two studies. Another explanation is that the nature of medical practice has changed. The lower AE rates may be due to medical innovations improving the safety of medical practice, while the higher negligence rates may indicate that physicians in New York less frequently met the standard of care than did those in California. Another and

related explanation is that we studied different populations. Because ours was a random sample, we may have studied hospitals providing higher percentages of suboptimal care; these kinds of hospitals may have opted out of the California sample because it was a convenience sample. Still another explanation is that our double physician review process led to a different estimate than did the single physician review used in the CMA study. All of these factors and others may be playing roles.

B. Sensitivity to Process and Design Factors

We sought to estimate all AEs that occurred in New York hospitals. From the record review, however, we would not find events that could be identified only after discharge unless they were described in notes in the hospital record sampled or they resulted in rehospitalization in New York. Most events that did not lead to rehospitalization were probably mild. Rehospitalization outside the state would lead to underestimation. Events that occurred outside New York and led to hospitalization within the state would cause overestimation of rates. We have no means of estimating the relative influence of these factors, but the effect of any imbalance is probably small, and we have ignored it.

Absence of relevant information in the hospital record may also have caused errors. Lack of information would lead to underestimation of AEs if the event was not identifiable or overestimation if mitigating circumstances were not recorded. In the intermediate pilot study we explored the relative importance of such possible discrepancies. Examination of litigation files and hospital risk management programs enabled us to identify those AEs that were not noted in the hospital record. Of 116 AEs identified in such files, in only 4 (3.4%) cases was the AE not discoverable because information had not been recorded in the hospital record. Similarly, of the 67 cases found by risk

management procedures to have evidence of negligence, 7 (10.4%) cases had inadequate data in the record to justify such a judgment by our physician-reviewers. These results suggest that undercounting due to lack of information in the record was low.

Another problem could be screening errors. In the intermediate pilot study of records known to have a high incidence of AEs, the false-negative rate was 7.9% at the screening stage for records containing an AE. Similarly, 6.3% of negligent cases were missed because of screening errors. These false-negative results would cause our sample estimates to be too low. In the intermediate pilot study, 28.2% of negative records were mistakenly screened as positive. The bulk of positive screening errors would be eliminated in the physician-review step, but the remainder may be counted among the physician false positives, causing our sample estimates to be too high.

At the final step, the physician-reviewer may have failed to uncover an AE for which there was evidence in the medical record or, on the other hand, mistakenly identified an event which was not attributable to medical management. In our intermediate pilot study, 14 (12.6%) of all AEs with clear evidence in the hospital record were missed by physician review. Reviewers failed to identify as negligent 8.9% of records with clear evidence of negligence. On the other hand, reviewers reported AEs in 15.4% (6/39) of cases in which a review of medical records supplemented by litigation records (the gold standard for purposes of this study) did not support such a finding, and negligence in 14.8% of cases (13/88) in which according to the gold standard, there was none. Thus, overcalls occurred more frequently than undercalls. The proportion of overcalls may be smaller in our population study, however, because we employed two physician reviews. In addition, because the intermediate pilot study focussed on records with a very high incidence of potential AEs, there was probably an increased proportion of marginally possible but not definite AEs. As a result, we believe the actual overcall rate

in the New York State sample is smaller than the intermediate pilot figures suggest.

In summary, our estimates of potential errors in the sampling process are small. Overcounting and undercounting errors appear generally to be of the same order of magnitude and thus seem not to bias the sample estimate in a particular direction.[1] Finally, to the extent that we can quantify them, all of our errors are within our 95% confidence estimates. For all of these reasons we use the figures derived from our sample survey to project incidence for the population at large.

C. Sensitivity of Estimate to Missing Records

The 4% of records not found in our first visit to the hospitals may have had a different rate of AEs than the 96% that were evaluated. To address this issue after the initial record review was completed, we asked all hospitals to verify the status of all 1,234 records missed on the first review. Of these 1,234 records, a total of 578 (47%) were subsequently made available to us for screening. Of the remaining 656 records (2.1% of the original sample) that were never found, 90 (0.3% of the total sample) were SPARCS mismatches: no hospital record could be found corresponding to SPARCS information. An additional 470 of these 656 records (1.5% of the total sample) were declared by record rooms to be lost. Another 96 (0.3%) were found but not available for review on second follow-up; 74 of these were in one hospital that had initiated a new record-keeping system and could not match new numbers with the hard copy. The remaining 22 had a miscellaneous set of problems, including record "burned in fire," "out for microfiche," or "located, but index hospitalization record missing."

[1] To determine the cumulative effect of errors in both directions requires making assumptions about the very incidence rates we seek to estimate, defeating the objectivity of adjustments based on these errors.

At six randomly selected hospitals physicians reviewed all records that were not available on our first visit but were subsequently located. These six hospitals had a total of 5,455 records in the original sample of 31,429. Of these, 326 (6.0%) were missing on first review. The percentage missing on first review varied across the six hospitals from a high of 20.4% to a low of 2.5%. Because we sampled hospitals for follow-up randomly with probability proportional to the number of missing records, the hospitals included in the six hospital follow-up have higher than average numbers of records missing at first review.

A total of 154 of the 326 missing records (47.2%) were found and reviewed at follow-up; the percentage of records found at the six hospitals ranged from 88% to 29%. These figures are comparable to the percentage of records located at the other hospitals on follow-up. (As noted in Chapter 5, we followed up on the status of missing records at <u>all</u> study hospitals.) Among the 172 records not located at these six hospitals, 151 were declared lost, 17 were SPARCS mismatches, and 4 did not include material concerning the index hospitalization.

This left the total of 154 records reviewed at the six hospitals; of these, 65 (42.2%) passed the screen by the medical record administrators for criteria met. In 8 of these 65, physician review determined the presence of an AE. Table 6.2 presents results of the review found on follow-up at these six hospitals.

Table 6.2

AE and Negligence Rates in Initial Review and Missing Record Review at Six Randomly Selected Hospitals

Hospital	AE Rates % Initial Visit	AE Rates % Follow-up[1] Visit	Negligence Rates % Initial Visit	Negligence Rates % Follow-up Visit
1	3.49	0.0	0.61	0.0
2	3.59	5.63	1.91	0.0
3	2.88	6.52	0.52	0.0
4	2.29	1.83	1.18	1.83
5	4.96	6.25	2.01	0.0
6	3.77	0.0	0.35	0.0
Weighted Totals	3.64	2.54	1.10	0.66

[1] To help protect the identity of the hospitals, we do not list the number of records on which these rates are based. The rates for initial visits are based on at least 500 records. For follow-up, the number of records reviewed ranged from 14 to 53. Therefore, the corresponding standard errors are large.

The AE rates for the follow-up records are lower than those found on the initial visit at three hospitals and higher in the other three. Because of the small numbers involved, none of these differences in individual hospitals approaches statistical significance. The overall six-hospital AE rate is lower for follow-up than for initial visit. The negligence rates are lower in five of six hospitals in the follow-up group in every instance but are based on too few cases to be meaningful.

The totals in Table 6.2 reflect the six-hospital average. Since these hospitals were randomly selected for follow-up, we can extrapolate their rates to form an estimate of the AE rates among all missing records at all 51 hospitals. With this extrapolation (weighting inversely by the probability of selection for follow-up), we calculate an AE rate of 3.55% over all for the missing records. This figure is higher than the six-hospital average of 2.54% because the highest follow-up AE rates were from the hospitals with the smallest number of missing records. The AE rate was 3.63% for the records found on initial

visits to all 51 facilities. This figure differs slightly from the 3.69% reported above because it does not adjust for missing records. From these results, we conclude that our results do not understate incidence because of missing records.

Since we found only about half of the 4% originally missing records, one might argue that our AE rate of 3.55% applies only to this fraction and that the 2.1% never located records could have had a much higher AE rate. We think this unlikely since 0.6% were due to SPARCS mismatches or other documented causes, which could be considered random, leaving only 1.5% that can be called lost. We compared the records originally found, those found on follow-up, and those still missing using information from the SPARCS data set. The three groups were similar with respect to age, gender and DRGs. Some evidence suggests that records found at follow-up had a higher percentage of blacks and Medicaid patients than those found at first, but there appears to be no real difference in those found at follow-up and those still missing. In the first survey 16% of the records were for black patients; the corresponding figures for follow-up and still missing records were 29% and 35%, respectively. The proportion of patients who were Medicaid recipients was 14% in the first survey, 22% in the follow-up, and 27% in those still missing.

We thus have substantial evidence that the missing records were not different from those included in this estimate. However, even if we assume the 3.7% over all AE rate applies to the population of records found either initially or on follow-up (98%), and that the rate among the never-found was triple that (11.1%), our over all rate would only be 3.8% (0.98 x 3.7 + 0.02 x 11.1). Thus, we conclude that missing records could have imparted no substantial bias to our results, and we made no adjustments in our estimates to account for them.

D. Sensitivity of Estimates to Range of Judgments Regarding Causation and Negligence

As noted in the previous chapter, not all physician judgments about causation and negligence can be unequivocally yes or no. Thus, we expected to find a broad range of ratings of confidence in the causal judgments; this was indeed the case. Tables 6.3 and 6.4 show the average of the two reviewers' scores of confidence in causation and negligence as well as the estimates of how these levels of confidence would be distributed if all records in the state had been reviewed. Chapter 5 describes in detail how these scores were obtained. A score of 0 means no evidence of causation. A score of 1 means some evidence but "little or no confidence" for a judgment of causation, and a score of 6 means "virtual certainty" in the judgment for causation. The scores of 3 and 4 are close calls. A score of 3 means "not quite likely, less than 50-50" and 4 means "more likely than not, more than 50-50."

The range of causal judgments, especially the proportion that are close calls, are important in estimating the costs of any administrative compensation plan. If most causal judgments are clustered in a grey zone of limited confidence in judgment, or if most were close calls, the administrative costs of any compensation plan would be great. On the other hand, if many of the determinations were either definitely positive or definitely negative for causation or negligence, administrative costs would shrink.

With respect to the distribution of causation scores, in the overwhelming majority of records reviewed, we found no evidence for an AE (i.e., a causation or AE score of 0). We estimate that 95% of the discharges in New York State were in this category (Table 6.3). The proportion of difficult causation determinations, those lying near the midpoint of confidence in causation scale, is small. Only 1.3% were in the 3 to 4 AE

score range (.34 + .41 + .52), or close-call AEs. For the state as a whole, however, the estimated number of AEs for this range of scores is 33,857.

Table 6.3
Range of Scores for Confidence in Judgments of Adverse Events

Average Score	Number in Sample	Population Estimates Percent	Number
0	28,617	95.05	2,539,701
1.0	36	0.11	2,938
1.5	16	0.04	1,027
2.0	33	0.12	3,103
2.5	62	0.25	6,652
3.0	95	0.34	8,962
3.5	129	0.41	10,871
4.0	166	0.52	14,024
4.5	184	0.62	16,715
5.0	255	0.87	23,336
5.5	292	0.95	25,452
6.0	236	0.71	19,082

The distribution of causal scores also demonstrates that the population estimate of AEs is sensitive to the definition of an AE. If we lower our threshold to include all cases with even minimal evidence of causation, ourestimated number of hospitalizations that led to AEs would increase. With this lower boundary, 1.2% of hospitalizations would be low-threshold AEs or have a score of 1.5-3.5. If we included the entire group of cases with scores of 1.5 or more, therefore, our estimate of the rate of AEs would increase by only 34%.

A minority of AEs were considered negligent. The number of close-call negligent cases was proportionately greater but absolutely smaller than the number of close-call AEs. The estimated number of close calls for score 3-4 negligence is 17,238, approximately half the number for close-call AEs (Table

6.4), but the distribution of confidence scores differs from that for AEs. Physician-reviewers judged a smaller proportion of negligent events with high confidence and a larger proportion in the low threshold range. With 1.0% scoring between 1.5 and 3.5, the estimated rate of negligent AEs would approximately double if the low-threshold cases were included.[2] This means that in many cases there was some evidence of negligence but not enough to reach our threshold finding and thus be characterized as negligent. It reinforces a point made in the previous chapter that AEs lie on a spectrum ranging from some with no evidence of negligence to those with overwhelming evidence of negligence. It also shows that the percentage of negligent AEs is very sensitive to the definition of negligence. Adding all the low-threshold negligent AEs would more than double the total number of negligent AEs. Conversely, if we required unequivocal proof of negligence (score > 5.0), the estimate of negligent AEs would diminish by three quarters.

[2]The reader should also note that part of the difference in distributions of AE scores and negligence scores arises from differences in our study protocols for judging causation and negligence. If physician-reviewer #1 found an AE (score equal to or greater than 1), but physician-reviewer #2 found none (score = 0), a senior physician performed an independent review (see Chapter 5) and determined an AE score equal either to 0 or greater than 1. The odd review was dropped from further analysis. Agreement was thus forced, resulting either in 2 scores of 0 (overall AE score = 0), or 2 scores in the range 1-6. For negligence determinations, in no case sent to senior physician review did two physicians differ on the confidence of negligence. Thus, if reviewer #1 found a confidence of negligence equal to 0, and reviewer #2 found a confidence of negligence of equal to or greater than 1, the two scores were averaged. This procedure left a greater chance for composite scores of negligence to fall in the range of 1-3.5.

Table 6.4
Range of Scores for Confidence in Judgments of Negligence

Average Score	Number in Sample	Population Estimates Percent	Number
0	29,368	97.54	2,606,121
0.5	49	0.14	3,766
1.0	76	0.23	6,197
1.5	45	0.14	3,628
2.0	106	0.30	8,162
2.5	73	0.24	6,386
3.0	70	0.22	5,909
3.5	54	0.17	4,516
4.0	67	0.26	6,813
4.5	64	0.22	6,023
5.0	77	0.27	7,337
5.5	47	0.17	4,503
6.0	25	0.09	2,502

These analyses provide some basis for estimating the relative feasibility of determining the merit of a claim for compensation in a no-fault system versus the merit of a claim for damages in a fault system. A no-fault system would focus solely on issues of causation.

E. Disabilities Caused by Adverse Events

AEs cause a wide span of effects from relatively minor to permanent disability and death (Table 6.5). To judge the severity of the AE in terms of its impact on the person's health, we asked physician-reviewers to estimate the disability caused by an AE, much as they might be asked to assess disability by reviewing the medical records of a person applying for disability benefits. (However, our physicians did not have the benefits of

interviewing the patient or performing a physical examination.) We then used these estimates to develop population estimates of disabilities (see Technical Appendix 5.VI.1).

Table 6.5

Estimated Population Distribution of Adverse Events in Disability Categories

Disability Category	Adverse Events Number (% ± SE)	Negligent Adverse Events Number (% ± SE)	Percent AEs with Negligence % ± (SE)
Minimal impairment, recovery 1 month	56,042 (56.8 ± 1.6)	12,428 (45.7 ± 3.7)	22.2 ± 2.8
Moderate impairment, recovery 1-6 months	13,521 (13.7 ± 1.1)	3,302 (12.1 ± 2.2)	24.4 ± 4.8
Moderate impairment, recovery > 6 months	2,762 (2.8 ± 0.5)	817 (3.0 ± 1.0)	29.6 ± 8.6
Permanent impairment, 1-50% disability	3,807 (3.9 ± 0.6)	869 (3.2 ± 1.1)	22.8 ± 6.8
Permanent impairment, >50% disability	2,550 (2.6 ± 0.4)	877 (3.2 ± 0.8)	34.4 ± 8.1
Death	13,451 (13.6 ± 1.7)	6,895 (25.4 ± 4.2)	51.3 ± 6.9
Cannot reasonably judge disability	6,477 (6.6 ± 0.7)	1,989 (7.3 ± 1.3)	30.7 ± 5.9
Total*	98,610 (100)	27,177 (100)	27.6 ± 2.4 p< .0001

*Totals differ from sums of those reported above because of rounding error.

Most AEs (56.8 ± 1.6%) resulted in what we defined as minimal impairment with complete recovery in one month (Class One). Permanent impairment with less than 50% disability (Class Four) resulted from 3.9% (± 0.6%) of AEs. Permanent total disability (Class Five) resulted from 2.6% (± 0.4%) of AEs, and deaths (Class Six) from 13.6% (± 1.7%). If we extrapolate these results to the State of New York in 1984, 2,550 individuals suffered permanent total disability as a result of AEs and 13,451

died. (As pointed out in Chapter 5, we made no effort to identify what was undoubtedly a substantial fraction of this group, namely, deaths in patients with serious underlying diseases and, therefore, limited life expectancy independent of the AEs.)

The proportion of AEs associated with negligence was not uniform across disability categories. The negligence rate was greater in patients who suffered more severe AEs. Of the estimated 56,042 AEs that led to disability lasting less than one month, only 12,428 (22.2 ± 2.8%) would be expected to be associated with negligence. On the other hand, of the 2,550 AEs that caused permanent total disability, 877 (34.4 ± 8.0%) would be expected to be associated with negligence. In addition, 51.3% (± 6.9%) of the 13,451 deaths from AEs would be attributable to negligence. These differences in the proportion negligent by category are highly significant (Wald test X^2 = 21.04, $p<.0001$). The increased proportion of AEs associated with negligence in the more severe disability category is consistent with the results of both the HEW study and the California Medical Association study (see Chapter 3).

III. Risk Factors for Adverse Events

A major question not addressed in previous studies is the distribution of AEs across different patient populations. Some might argue that AEs occur infrequently and are largely out of the control of health care management. Others would argue certain groups are at greater risk for AEs. We have tried to identify individual and hospital factors that might be associated with higher rates of AEs or negligent care.

A. Individual Factors

1. Age. We first analyzed the possible role of age as a risk factor for AEs (Table 6.6). The rates shown are weighted by the inverse of the sampling probability so as to reflect rates for the total population (see Technical Appendix 5.V.2). The standard errors are computed controlling for the sampling design as explained in Technical Appendix 5.V.3. Note the substantial (nearly tenfold) increase in the AE rate with increasing age.

Table 6.6

Adverse Event Rates in New York by Age

Age Group	# of Cases Reviewed[a]	Percent Adverse Event[b]	Standard Error
Newborn	3,595	0.6	0.1
15 yrs. or less	3,066	2.1	0.4
16-44	11,101	2.6	0.2
45-64	7,379	4.7	0.4
65+	4,980	5.9	0.5
Totals	30,121	3.7	0.2

a = unweighted
b = weighted

The reasons for hospitalization are likely to be different for different age groups. Therefore, some of the increase shown in Table 6.6 may be due to the more complex conditions common in older people. In order to control for this, we standardized the rates for four DRG groups (see Technical Appendix 5.V.4). The DRG standardization was based on estimates of the probability that an AE would occur in a patient with a given DRG simply as a result of the complexity of care required. For example, an AE would be much more likely in a patient undergoing removal of a brain tumor than in one whose superficial skin abscess was being drained (see Technical Appendix 5.V.4). Such standardization controls for the differences in underlying disease states that

occur among age groups or other groups we might analyze.

The spread of variations in AE rates across age groups is less in the standardized figures shown in Table 6.7, indicating that as expected that some of the difference in rates reflected differences in complexity of the underlying diseases. But there remains a difference in the likelihood of an AE across the range of ages. Newborns have a relatively low risk of suffering an AE. The differences in AE rates increased with age and are statistically highly significant. Thus, an AE cannot be considered a purely random event, and increasing age is clearly a risk factor.

Table 6.7

Adverse Event Rates in New York State by Age
Standardized by DRG class

Age Group	Rate	Standard Error
Newborn	1.4	0.3
15 or less	2.7	0.6
16-44	2.6	0.2
45-64	4.4	0.4
65+	5.7	0.6
Totals	3.7 p < .0001	0.3

Table 6.8 describes the relationship of negligent AEs to age. Because the denominator of every percentage of negligence figure is the number of AEs, and since more complex care leads to more AEs, the calculations implicitly control for the complexity of care in the percentage of negligence calculation. Note the statistically significant relationship between percentage of negligent AEs and age. As shown in Table 6.8, this is due almost entirely to a much higher incidence rate among people over age 65.

Table 6.8

Percentage of Negligent Adverse Events in New York State by Age

Age	Percent of Adverse Events Found Negligent	Standard Error
Newborn	20.8	7.1
15 or less	21.9	6.0
16-44	26.7	2.8
45-64	20.6	2.4
65+	33.1	4.2
Totals	27.6	2.6
p value	p = .01	

Because the differences among the under-65 groups were not statistically significant, we standardized the percentage of negligence on the basis of only two age categories: 65 years or older and 64 years and younger. We did not standardize the percentage of negligence using DRG levels because that control is implicit when we calculate the percentage of AEs that are negligent (Chapter 5, Section VI D). Probably for that reason we found that the percentage of negligent AEs did not vary by DRG level. That finding was also expected because we asked reviewers to take into account the complexity of the case when judging negligence. Thus, differences in complexity of care, as represented by DRG levels, should not, and do not, affect the estimated percentage of negligent events.

Table 6.9 summarizes the information from the previous three tables. This is the format we also use in subsequent tables, each of which includes the concepts of standardization noted above.

Table 6.9
Rates of Adverse Events and Negligence by Age

Age	# Cases Reviewed	Adverse Events Crude %	Crude SE	Standardized[1] %	Standardized SE	Negligence %	Negligence SE
Newborn	3,595	0.6	0.1	1.4	0.3	20.8	7.1
0-15	3,066	2.1	0.4	2.7	0.6	21.9	6.0
16-44	11,101	2.6	0.2	2.6	0.2	26.7	2.8
45-65	7,379	4.7	0.4	4.4	0.4	20.6	2.4
6>65	4,980	5.9	0.5	5.7	0.6	33.1	4.2
p value				p = .0001		p = .01	

1 standardized for DRG class

2. Gender. As Table 6.10 shows, we found no significant differences between genders in either AE rates or percentage of AEs associated with negligence. We expected this to be the case.

Table 6.10
Adverse Event and Negligence Rates by Gender

Gender	#Cases Reviewed	Adverse Events Crude %	Crude SE	Standardized[1] %	Standardized SE	Negligence Crude %	Crude SE	Standardized[2] %	Standardized SE
Male	13,722	3.83	0.32	3.75	0.41	28.3	2.9	27.4	2.8
Female	16,399	3.60	0.25	3.68	0.36	27.0	3.2	25.0	2.8
p value		p = .87						p = .48	

[1]Standardized for age and DRG class
[2]Standardized for age (<65, 65+)

3. Race. Although we are not aware of previous inquiries seeking possible relationships between AEs and negligence and social groups, others have described differences in care provided various racial groups. For example, Yergan and colleagues demonstrated that nonwhite patients with pneumonia received fewer

hospital services than expected on the basis of their medical problems.[3] Bombardier and colleagues studied eleven noncardiac surgical procedures and found that whites underwent significantly more procedures than blacks.[4] When they controlled for income, however, the interracial differences became smaller and were no longer statistically significant. Recently, Wenneker and Epstein found less coronary angiography, coronary artery bypass grafting and coronary angioplasty among blacks.[5] There is also evidence that blacks undergoing cardiac surgery are more likely to receive care from a physician in training than from a senior staff physician.[6]

Lesser utilization of resources does not necessarily mean that blacks receive poorer medical care, but a series of studies has demonstrated that neonatal mortality is elevated in black children at all income levels.[7] Another study has suggested more preventable deaths in blacks than in whites in Alameda County, CA.[8] Nonwhite patients have only half the chance of receiving a renal transplant compared to white patients of the same age and sex.[9]

We did not find statistically significant racial

---

[3] Yergan, Flood, Logerfo & Diehr, Relationship Between Patient Race and the Intensity of Hospital Services, 25 Med. Care 592-603 (1987)

[4] Bombardier, Fuchs, Lillard et al, Social Economic Factors Affecting the Utilization of Surgical Operations, 297 N. Engl. J. Med. 699-705 (1977)

[5] See Wenneker & Epstein, Racial Inequalities in the Use of Procedures for Patients with Ischemic Heart Disease in Massachusetts, 261 J.A.M.A. 253-57 (1989)

[6] See Egbert & Rothman, Relation Between Race and Economic Status of Patients and Who Performs Their Surgery, 297 N. Engl. J. Med. 90-91 (1977)

[7] See Wise, Kotelchuck, Wilson & Mills, Racial and Socioeconomic Disparities in Childhood Mortality in Boston, 313 N. Engl. J. Med. 360-6 (1985); Lammer, Brown, Anderka & Guyer, Classification and Analysis of Fetal Deaths in Massachusetts, 261 J.A.M.A. 1757-62 (1989); Griffiths, White & Stonehouse, Ethnic Differences in Birth Statistics from Central Birmingham, 298 Br. Med. J. 94-95 (1989)

[8] See Woolhandler, Himmelstein, Silber et al, 15 Int. J. Health Services 1 (1985)

[9] See Kjellstrand, Age, Sex and Race Inequality in Renal Transplantation, 148 Arch. Intern. Med. 1305-09 (1988)

differences in AE rates and percentage of AEs due to negligence. The crude rates of AEs were highest for whites and lowest for Hispanics (Table 6.11). In our sample, however, the proportion of blacks in the older age groups is much lower than the corresponding proportion for whites. As noted in the discussion, other studies have found that blacks have fewer complex procedures and therefore would be expected to have a lower risk of an AE. We standardized by age and DRG group to obtain comparable rates for each racial group. The effect of standardization was to lower slightly the AE rate for whites and to increase the AE rates for blacks and Hispanics.

As for all comparisons in this chapter, the primary statistical test is for any difference among groups. This test was not significant after standardization. The standardized AE rates and the percentage of AEs due to negligence were, nonetheless, substantially higher for blacks than whites. The difference in a two-way comparison of percentage of negligence between blacks and whites was nominally significant ($p < .05$). This result has only suggestive value, however, because the comparison was made of the extreme groups after examining differences among three racial classifications (Table 6.11).

Table 6.11

Adverse Events and Negligence Rates by Race

Race	# Cases Reviewed	Adverse Events Crude %	Crude SE	Adverse Events Standardized[1] %	Standardized[1] SE	Negligence Crude %	Crude SE	Negligence Standardized[2] %	Standardized[2] SE
Whites, others	23,572	3.8	0.3	3.6	0.3	25.9	2.9	24.0	2.3
Blacks	4,808	3.5	0.5	4.8	1.0	37.2	5.8	37.9	6.3
Hispanics	1,741	2.5	0.5	3.4	1.1	25.9	4.9	29.4	8.6
p value				p = .48				p = .12	

[1] standardized for age and DRG class
[2] standardized for age (<65, 65+)

4. Payer Status. Payer status is a proxy for socioeconomic class. As such it may be confounded with race because poverty levels are higher in the nonwhite population. Indeed, ample evidence suggests that uninsured individuals, many of whom are indigent, have decreased access to care. Further, the care they receive may be of poorer quality.

For example, individuals without health insurance receive less primary and preventive care than do others.[10] Lurie et al. found that patients whose Medi-Cal was terminated because of cost containment initiatives in California in the early 1980s had less treatment of high blood pressure and of diabetes.[11] Braveman et al. found a significant increase in AEs among pregnant women without insurance.[12]

Although payer status does not provide the information regarding socioeconomic status obtained from data regarding income levels and education, it is often used in health services research as a proxy for it. We created four payer categories: self-pay, Medicare, Medicaid, and Blue Cross or other commercial insurance. We analyzed payer class in two segments. Because most individuals who receive Medicare are over 65, and a simple analysis of payer status would be confounded by age, we first excluded Medicare recipients (Table 6.12).

---

[10] For instance, Woolhandler and Himmelstein demonstrated that uninsured individuals have less screening for high blood pressure, cervical and breast cancer, and glaucoma testing. See Woolhandler, S., Himmelstein, D., Reverse Targeting of Preventive Care Due to Lack of Health Insurance, 259 J.A.M.A. 2872 (988). These results are supported by other information from the Rand Health Insurance Experiments that showed that out-of-pocket co-payments reduced the use of preventive care. See Lurie, N., Manning, W., Peterson, C., et al., Preventive Care: Do We Practice What We Preach? 257 J.A.M.A. 801 (1987).

[11] See Lurie, Ward, Shapiro & Brook, Termination from Medical: Does It Affect Health? 311 N. Engl. J. Med. 480 (1984)

[12] See Braveman. Oliva, Miller et al, Lack of Health Insurance and Adverse Outcomes in Newborns in an Eight County Area of California, 1982-86, N. Engl. J. Med., in press

Table 6.14

Adverse Event and Negligence Rates in Hospitals with Differences in Minority Discharges

Percent Minority	# Cases Reviewed	Adverse Events Crude %	Adverse Events Crude SE	Adverse Events Standardized[1] %	Adverse Events Standardized[1] SE	Negligence Crude %	Negligence Crude SE	Negligence Standardized[2] %	Negligence Standardized[2] SE
> 83	3,107	3.14	0.37	4.92	1.0	38.8	4.3	37.0	4.5
15-83	6,382	2.30	0.42	2.49	0.54	31.2	6.4	29.1	6.3
< 15	20,632	2.52	0.24	2.38	0.29	22.8	3.4	21.5	3.0
p value				p = .048				p = .01	

[1] standardized for age and DRG
[2] Standardized for age (<65, 65+)

When we standardized for DRG and age, the differences in AE rates between hospitals with predominantly minority populations and others widened. These differences were clearly statistically significant. In addition, the percentage of AEs due to negligence increased significantly as the percentage of minorities increased. These findings suggest that differences between racial groups in AEs rates and negligence reflected, at least in part, differences between hospitals.

This conclusion is buttressed by the fact that rates of AEs and negligent AEs among blacks and whites were not significantly different at those hospitals that provide care for a mix of races similar to the population at large. Table 6.15 shows the rates of AEs and percentage of negligent AEs by racial group when we excluded those hospitals serving predominantly minority people or only white people.

Table 6.15

Adverse Event and Negligence Rates Among Racial Groups in Hospitals Serving White and Minority Patients

Race	# Cases Reviewed	Adverse Events Crude %	Crude SE	Standardized[1] %	Standardized[1] SE	Negligence Crude %	Crude SE	Standardized[2] %	Standardized[2] SE
Whites	22,714	3.91	.28	3.69	3.3	25.9	2.9	24.2	2.5
Black	2,854	2.96	.58	3.65	1.02	24.2	4.9	24.2	5.9
Hispanic	813	1.96	.48	2.80	1.5	9.3	3.5	9.4	4.7
p value				p = .84				p = .007	

[1] Standardized for age and DRG class
[2] Standardized for age (<65, 65+)

The standardized AE rates and percentage of negligent rates were virtually identical for blacks and whites in those hospitals that serve a racially mixed population.

In conclusion, there were differences in AE and negligent AE rates between racial and payer groups but none were significant. However, significant racial differences were revealed when controlling for payer status. More importantly, there were significant differences between hospitals that serve a predominantly minority population and other hospitals. That is, blacks were more likely to be hospitalized at institutions with more AEs and higher rates of negligence. We now turn to a further analysis of hospital characteristics.

B. Hospital Factors

As noted in Chapter 5, we believe that while AE rates may be independent of the quality of care, the percentage of AEs due to negligence is a measure of quality. Therefore, we believe our study provides one way to assess quality of care. Much of the recent focus on interhospital variations in quality of care has rested on standardized mortality data. The latter permit

comparisons of hospitals, especially when the rates are adjusted for clinical characteristics of the hospital case mix.[13] However, there can be serious problems interpreting differences in mortality rate, and at present studies of this subject are not extensive.[14] Researchers interested in exploring the differences in aspects of care provided by types of hospitals have selected a number of parameters to study. Several investigators have explored the differences in the economic performances of nonprofit and for-profit institutions.[15] Others have looked at the differences between teaching and non-teaching hospitals[16] and between rural with urban hospitals.[17] Also of interest have been differences between public or municipal hospitals and other nonprofit hospitals.[18] Finally, investigators have examined the attributes of different size hospitals, measured by number of discharges per year.[19]

Very few studies have analyzed the differences in quality of care rendered in different types of hospitals except for those

---

13 See Dubois, Brook & Rogers, Adjusted Hospital Death Rates: A Potential Screen for Quality of Medical Care, 77 Am. J. Public Health 1162-66 (1987)

14 Fink, Yano & Brook, The Condition of the Literature on Differences in Hospital Mortality, 27 Med. Care 315-16 (1989)

15 See Renn, Schramm, Watt & Derzon, The Effects of Ownership and System Affiliation on the Economic Performance of Hospitals, 22 Inquiry 219-36 (1985); Sloan & Vraciu, Investor Owned and Not for Profit Hospitals: Addressing some issues, 2 Health Affairs 25-38 (1983)

16 See Becker & Sloan, Utilization of Hospital Services: The Roles of Teaching, Case Mix and Reimbursement, 20 Inquiry 248-57 (1983); Richards, Lurie, Rogers & Brook, Measuring Differences Between Teaching and Non-teaching Hospitals, 26 Med. Care Supplement 1-141 (1988); Garber, Fuchs & Silverman, Case Mix, Cost and Outcomes, Differences Between Faculty and Community Services in a University Hospital, 310 N. Engl. J. Med. 1231-37 (1984)

17 See Moscovice, Rural Hospitals: A Literature Synthesis in Health Services Research Agenda, 23 Health Serv. Res. 891-930 (1989); Hein, The Quality of Perinatal Care in Small Rural Hospitals, 240 J.A.M.A. 2070-72 (1978).

18 See Tetelman, Public Hospitals -- Critical or Recovering? 88 Health Serv. Rep. 295-304 (1973)

19 See Miransky, Kerner, Sturgeon, Zauber et al, A Comparison of Primary Breast Cancer Management in Small, Intermediate, and Large Community Hospitals and a Comprehensive Cancer Center, 156 Progr. Clin. Biol. Res. 87-96 (1984)

based on standardized mortality data.[20] [21] [22] The Teamsters Union study of the early 1960s emphasized the importance of medical staff adherence to clinical protocols as a determinant of variations in the quality of care rendered at different hospitals.[23] Among those characteristics of hospital medical staffs which appear to influence quality, Shortell found the most important to be organizational structures which promote physician involvement in quality review.[24] Flood and Scott compared the quality of care rendered at various institutions with different organizational arrangements.[25] Bernard-Stevens compared methods for monitoring quality of care at rural and urban medical centers.[26] These studies provide at best only tentative data regarding variations in the quality of care.[27]

Because we selected a random sample of 51 New York hospitals, they served as a base for analyses of variations in care. We chose to describe variation among hospitals on four dimensions. First, we looked at variations by hospital size

---

[20] For two excellent reviews of this literature, see Donabedian, The Epidemiology of Quality, 22 Inquiry 282-92 (1985); Palmer & Reilly, Individual and Institutional Variables Which May Serve as Indicators of Quality of Medical Care, 17 Med. Care 693 (1979)

[21] See Kosecoff, Fink, Brook et al, General Medical Care and the Education of Internists in University Hospitals. An Evaluation of the Teaching Hospital General Medicine Group Practice Program, 102 Ann. Intern. Med. 250-57 (1985)

[22] See Feigenson, Feigenson, Gitlow et al, Outcome and Cost for Stroke Patients in Academic and Community Hospitals, A Comparison of Two Groups Referred to a Regional Rehabilitation Center. 240 J.A.M.A. 1878-80 (1978)

[23] M.A. Morehead et al, A Study of Hospital Care Secured by a Sample of Teamster Family Members in New York City, New York, Columbia School of Public Health and Administrative Medicine (1964)

[24] See Shortell & Lofergo, Hospital Medical Staff Organization and Quality of Care: Results for Myocardial Infarction and Appendectomy, 19 Med. Care 1041 (1981)

[25] See, generally, A.B. Flood & W.R. Scott, Hospital Structure and Performance, Baltimore, Johns Hopkins Press (1987)

[26] See Bernard-Stevens, Gust, Moore & Zetterman, Monitoring Quality of Care in Urban and Rural Medical Centers in Nebraska, 67 Nebr. Med. J. 124-26 (1982)

[27] There has been, however, a great deal of research regarding the importance of volume of procedures performed at hospitals and quality of care. See e.g., Hughes, Hunt & Luft, Effects of Surgeon Volume and Hospital Volume On Quality of Care in Hospitals, 25 Med. Care 489-503 (1987).

based on number of discharges in 1984. Next, we analyzed the role of ownership, comparing nonprofit, government, and proprietary hospitals. Third, we analyzed differences in hospitals in seven locations, clustering hospitals in New York City, Nassau, Suffolk, Westchester counties, upstate metropolitan areas, and upstate non-metropolitan statistical (MSA) hospitals. Fourth, we developed information on hospital teaching status. We developed a method for defining teaching status based largely on the New York State Department of Health's SEDI Index (see Technical Appendix 4.II.1). Using the Index, we constructed categories including university teaching hospitals, affiliated teaching hospitals, and non-teaching hospitals.

As discussed in Chapter 4, our sample was designed to include discharges from the full spectrum of hospitals in New York State. To analyze differences among hospitals, we included only those events known to have resulted from management during the index hospitalization. Referring back to Table 6.1, which shows the relation of management to discovery for the AEs, we consider here only the first three categories:

1. The 647 events occurring and discovered in the index hospitalization;
2. the 78 events occurring in the index hospitalization and discovered in subsequent outpatient care; and
3. the 67 events occurring in the index hospitalization and discovered subsequently in inpatient care.

Those 792 events that were attributable with certainty to the index hospitalization were lower than the total figure reported in previous sections because we did not include cases in which the AE was discovered in a different institution. AEs known to have occurred in the sampled hospital totalled 792 (out of 30,121 cases), or a weighted rate of 2.54% $\pm$ 0.20%, over all. There were 182 negligent AEs; the over all percentage of AEs due to negligence was 26.3% $\pm$ 2.8%.

1. Individual Hospitals. We next computed AE rates for individual hospitals. Our sampling process involved linking any hospitals with shared administrative facilities, specialty hospitals with nearby general hospitals and small hospitals (fewer than 8,000 1984 discharges) with neighboring hospitals, to form hospital clusters. This yielded a sample of 31 hospital clusters, which included 51 separate facilities. Some of these facilities have the same administrative control. For purposes of presenting hospital-specific rates, all facilities under a single administrative control were considered to be the same hospital unit. This reduced our sample of 51 facilities to 44 hospital units. Six of these 44 units had fewer than 200 records in the sample. In displaying figures for individual hospitals, we excluded those hospitals with fewer than 200 records in the sample because of the instability of their rates, leaving 38 individual hospitals. All facilities are included in all other comparisons. Because of small numbers, we do not display negligence rates for individual hospitals.

The hospital-specific AE rates and standard errors for the 38 hospitals are shown in Table 6.16. Individual hospital rates ranged from 8.15% to 0.53% for AEs, a fifteenfold difference.

Table 6.16

Hospital-Specific Adverse Event Rates

Hospital	Crude Rate[1] ± SE	Age-DRG Standardized Rate[2] ± SE
1	8.15 ± 1.74	
2	6.10 ± 1.02	5.78 ± 3.78
3	5.22 ± 0.70	4.33 ± 2.24
4	5.14 ± 1.00	
5	4.80 ± 0.80	4.28 ± 2.13
6	4.49 ± 1.63	
7	3.53 ± 0.70	3.24 ± 1.77
8	3.46 ± 1.44	
9	3.39 ± 1.09	
10	3.37 ± 0.60	4.18 ± 2.46
11	3.29 ± 0.64	3.55 ± 2.55
12	3.28 ± 1.08	
13	3.11 ± 0.67	
14	2.94 ± 0.74	
15	2.34 ± 0.63	
16	2.31 ± 0.53	3.68 ± 3.07
17	2.30 ± 0.62	2.15 ± 1.51
18	2.24 ± 0.52	2.25 ± 1.56
19	2.14 ± 0.53	
20	2.02 ± 0.51	2.16 ± 1.72
21	1.90 ± 0.73	
22	1.87 ± 0.43	1.96 ± 1.47
23	1.86 ± 0.80	
24	1.82 ± 0.82	
25	1.76 ± 0.52	
26	1.71 ± 0.66	
27	1.67 ± 0.44	1.45 ± 1.52
28	1.59 ± 0.45	1.52 ± 1.14
29	1.59 ± 0.52	
30	1.56 ± 0.52	
31	1.45 ± 0.45	1.27 ± 1.00
32	1.36 ± 0.47	1.26 ± 1.39
33	1.26 ± 0.45	1.05 ± 0.95
34	1.18 ± 0.46	1.09 ± 1.06
35	1.17 ± 0.48	
36	0.86 ± 0.51	
37	0.75 ± 0.30	0.92 ± 1.15
38	0.53 ± 0.28	0.56 ± 0.83
Total[3]	2.54 ± 0.21	

[1] We exclude six hospitals with fewer than 200 records reviewed.
[2] Rates standardized for age and DRG are computed only for hospitals with at least 900 records in the sample.
[3] Total includes data from all 44 hospitals, but only events occurring in the index hospitalization.

Much of the variation in AE rates could be attributed to differences in patients' ages and diagnoses. Because of the relatively small numbers of records in many hospitals, hospital-specific rates standardized for age and DRG are not possible, but we do show such standardized rates for the 19 hospitals with at least 900 records in the sample. For this standardization, we combined the first two ages categories (newborn and children) to avoid sparse strata, leaving a total of 13 age by DRG strata. For 19 hospitals, the crude rates ranged from 0.53% to 6.10%; after standardizing for age and DRG, the range was from 0.56% to 5.78%. Thus, for the hospitals with at least 900 records in the sample, an elevenfold difference in crude rates decreased only slightly to a tenfold difference in standardized rates. Even though we restricted our standardization to hospitals with at least 900 records, the standard errors of the standardized rates are substantially higher than those of the crude rates.

2. Size of Hospital. First we looked at the variation among different sizes of hospitals. Table 6.17 presents the AE and negligence rates in hospitals grouped according to the total number of discharges in 1984. The standardized AE rate among hospitals with fewer than 8,000 discharges was about 60% of that for hospitals with more than 8,000 discharges. This result is only marginally significant ($p<.10$). The percentage of events due to negligence was unrelated to hospital size.

Table 6.17

Adverse Event and Negligence Rates in Hospitals with Different Numbers of Discharges

1984 Discharges	# Cases Reviewed[3]	Adverse Events Crude %	Adverse Events Crude SE	Adverse Events Standardized[1] %	Adverse Events Standardized[1] SE	Negligence Crude %	Negligence Crude SE	Negligence Standardized[2] %	Negligence Standardized[2] SE
< 8000	4,032 (16)	2.04	0.30	1.69	0.37	26.4	7.2	24.1	6.1
8000-19,999	16,332 (24)	2.67	0.30	2.71	0.42	27.3	3.5	25.2	3.2
≥ 20,000	9,757 (11)	2.64	0.39	2.84	0.48	24.7	5.6	24.1	5.5
p value				p = .08				p = .98	

[1] Standardized for age and DRG
[2] Standardized for age (<65, 65+)
[3] Number of facilities is given in parentheses

3. Hospital Ownership. Comparing hospitals according to ownership is complicated by the fact that the three proprietary hospitals in our sample had no newborns and very few discharges of patients under 16 years of age. To deal with this problem we did two comparisons. First, we computed standardized rates for nonprofit and government hospitals using all 17 age-DRG strata. We then dropped all discharges of persons younger than 16 and recomputed standardized rates for nonprofit, proprietary and government hospitals, using only 9 age-DRG strata. These comparisons are shown in Tables 6.18 and 6.19.

Table 6.18

Adverse Event and Negligence Rates
in Nonprofit and Government Hospitals

Ownership	# Cases Reviewed[3]	Adverse Events Crude %	Adverse Events Crude SE	Adverse Events Standardized[1] %	Adverse Events Standardized[1] SE	Negligence Crude %	Negligence Crude SE	Negligence Standardized[2] %	Negligence Standardized[2] SE
Nonprofit	24,226 (40)	2.61	0.25	2.61	0.30	25.5	3.4	23.6	2.9
Government	3,933 (8)	2.39	0.36	2.58	0.71	38.8	6.2	36.9	5.6
p value				p = .98				p = .03	

1 Standardized for age and DRG
2 Standardized for age (<65, 65+)
3 Number of facilities is given in parentheses

The rate of AEs differed little between the nonprofit and government institutions (Table 6.18), but the percentage of AEs due to negligence was nearly 50% higher in government hospitals compared to the nonprofits. This difference is statistically significant ($p<.05$).

Table 6.19

Adverse Event and Negligence Rates in Hospitals
of Different Ownership

Ownership	# Cases Reviewed[3]	Adverse Events Crude %	Adverse Events Crude SE	Adverse Events Standardized[1] %	Adverse Events Standardized[1] SE	Negligence Crude %	Negligence Crude SE	Negligence Standardized[2] %	Negligence Standardized[2] SE
Nonprofit	18,807 (40)	2.98	0.28	2.78	0.32	26.0	2.3	23.8	2.8
Proprietary	1,800 (3)	2.18	0.37	1.98	0.55	7.8	2.6	11.6	5.0
Government	2,853 (8)	2.85	0.46	2.86	0.78	40.5	6.6	38.1	6.2
p value				p = .42				p = .004	

1 Standardized for age and DRG
2 Standardized for age (16 - 65 years)
3 Number of facilities is given in parentheses

Excluding all patients younger than 16 years, we can compare rates in all three types of hospitals. The AE rates were not significantly different, but the percentages of negligent AEs were: that figure for proprietary hospitals was 11.6 (± 5.0), and for government hospitals, 38.1 (± 6.2), a greater than threefold difference.

4. Hospital Location. Differences in the rates of AEs by location of hospitals was particularly striking (Table 6.21). Non-MSA hospitals had extraordinarily low standardized AE rates (0.79%). Hospitals in New York City had very high standardized AE rates (3.45%). The percentages of AEs due to negligence were not significantly different. The low AE rate in non-MSA hospitals may reflect the fact that relatively few complex problems are managed there. Those procedures that involve higher risks of AEs are more likely to be undertaken at metropolitan tertiary centers.

Table 6.20

Adverse Event and Negligence Rates
by Hospital Location

Location	# Cases Reviewed[3]	Adverse Events Crude %	Crude SE	Standardized[1] %	Standardized[1] SE	Negligence Crude %	Crude SE	Standardized[2] %	Standardized[2] SE
NYC	11,748 (21)	3.28	0.46	3.45	0.57	23.1	4.5	21.6	3.9
Nassau-Suffolk-Westchester	6,594 (8)	1.98	0.16	1.96	0.31	38.4	8.7	34.9	7.1
Upstate Metropolitan (MSA)	9,849 (18)	2.41	0.29	2.32	0.37	25.9	3.4	25.1	3.4
Upstate Non-Metropolitan (non-MSA)	1,930 (4)	0.78	0.06	0.79	0.23	14.6	5.1	19.4	4.7
p value				p< .0001				p = .31	

[1] Standardized for age and DRG
[2] Standardized for age (<65, 65+)
[3] Number of facilities given in parentheses

5. Teaching Status

There were also striking differences in AE and negligence rates of hospitals compared on the basis of teaching status. The university teaching hospitals had higher AE rates, perhaps resulting from their more complex case mix, which was only partially controlled for by our standardization. Non-teaching hospitals, on the other hand, had low AE rates. When we looked at percentage of AE rates that were negligent, however, university teaching hospitals (11.3%) compare favorably with other hospitals, especially affiliated teaching programs (29.9%). We believe the high adverse rates in university teaching hospitals care reflects the more complicated medical problems in those institutions. We interpret the low negligent AE rates in those institutions to reflect higher quality care.

Table 6.21

Adverse Event and Negligence Rates in Teaching and Non-Teaching Hospitals

Teaching	# Cases Reviewed[3]	Adverse Events Crude %	Adverse Events Crude SE	Adverse Events Standardized[1] %	Adverse Events Standardized[1] SE	Negligence Crude %	Negligence Crude SE	Negligence Standardized[2] %	Negligence Standardized[2] SE
Non-teaching	14,100 (25)	1.97	0.16	1.79	0.21	27.7	3.7	24.4	3.1
Affiliate	12,034 (22)	2.88	0.36	3.35	0.53	31.6	3.8	29.9	3.7
University	3,987 (4)	3.86	0.61	3.66	0.76	10.9	2.8	11.3	2.8
p value			p = .003				p < .0001		

[1] Standardized for age and DRG
[2] Standardized for age (<65, 65+)
[3] Number of facilities in parentheses

In summary, many factors undoubtedly influence the likelihood that a patient will sustain iatrogenic injury. Complexity of treatment is almost certainly one. Complicated technology offers many more opportunities for things to go wrong.

The severity of a disease and the coexistence of more than one disease and a host of other patient factors, such as malnutrition and substance abuse, all influence the risk of AEs. Access problems may lead to late treatment and thereby increase risk. We believe these factors may explain much more of the variation in AE rates than does poor quality care.

On the other hand, our method of judging negligence expressly required reviewers to consider extenuating factors in assessing the standard of care rendered. Therefore, we believe the percentage of AEs that are negligent does represent a measure of quality of care.

University teaching hospitals, perhaps because they care for patients with more complicated problems, have high numbers of AEs, but the percentage of AEs due to negligence is low. On the other hand, affiliate teaching hospitals (that is, those with very limited residency programs), and municipal hospitals have much higher percentages of AEs due to negligence. Non-MSA hospitals have low rates of AEs, likely reflecting the less complex case mix. Proprietary hospitals have lower rates of negligent AEs. Hospital size did not affect AE rates or percentage of negligence rates.

IV. Types of Adverse Events

We now take a closer look at the AEs we have identified in terms of the types of injuries and the types of treatments they followed. Whether one chooses to compensate for medical injury through the tort system, by a no-fault compensation scheme, or by some other method, the most desirable solution to the problem clearly would be to prevent iatrogenic injury. Prevention of AEs is thus a fundamental issue for our study to address.

While in theory all AEs should ultimately be preventable, for many of them this will require a substantial increase in scientific knowledge and improvement in techniques of care.

Others, however, should be amenable to improvements in existing care systems. Clearly, one category of preventable AEs is that group due to negligence. These AEs result not from inadequate knowledge, but from failure to apply what is known, from failure of the physician or other caretaker to provide the accepted level of care. A second group of AEs that may be preventable at the current state of our knowledge are systems errors, AEs that result more from malfunctions of the organization of care than from individual shortcomings.

In both groups, the greatest impact from preventive efforts will be realized by reducing the number of AEs that cause serious lasting disability, level 3 or more on our scale. With these considerations in mind, we look at the types of AEs that occurred and note the percentage of each due to negligence and the fraction that resulted in serious disability.

AEs were classified by MPS physician-investigators after reading the descriptions of the AE written in the Adverse Event Analysis Form by both of the physician-reviewers for each case. Problem cases were discussed and classified by consensus. An AE was considered an operative complication if it occurred within the first two weeks following surgery, or if it was judged to be caused by an operation, regardless of when it occurred. See Chapter V for more details about the classification method.

A. Overview of Types of Adverse Events.

Table 6.22 lists the number of AEs in the sample and the estimated percentage of all AEs in New York in each category of AE. Nearly half (48%) of AEs were the consequence of an operation. Wound infections were the most common type of operative complication, accounting for 13.6% of all AEs. Acute technical complications (operative injury, bleeding, wound problems) were responsible for 13% of AEs. Postoperative bleeding was the most common technical complication but was

rarely life-threatening or a cause of prolonged disability. Noninfectious problems with incisions, such as wound separation and incisional hernia, also made up a sizable fraction of the technical complications of surgery. Late complications, such as vascular graft thrombosis, ureteral stricture, or retained common duct stone, accounted for 11% of AEs.

Table 6.22

Types of AEs and Negligent AEs

	Number of AEs in Sample	Percent of AEs in Population (weighted)	Percent of AEs due to Negligence (weighted)	Percent of AEs with Disability Severity >2 (weighted)
Operative				
Wound Infections	160	13.6	12.5	17.9
Technical Complications	157	12.9	17.6	12.0
Late Complications	137	10.6	13.6	35.7
Nontechnical	87	7.0	20.1	43.8
Failures	58	3.6	36.4	17.5
Subtotal, Operative AEs	599	47.7	17.0	24.0
Nonoperative				
Drug-related	178	19.4	17.7	14.1
Diagnostic Mishap	79	8.1	75.2	47.0
Therapeutic Mishap	62	7.5	76.8	35.4
Procedure-related	88	7.0	15.1	28.8
Falls	20	2.7	*	*
Fractures @	18	1.2	*	*
Postpartum @	18	1.1	*	*
Anesthesia	13	1.1	*	*
Neonatal	29	0.9	*	*
Other	29	3.3	*	*
Subtotal, Nonoperative AEs	534	52.3	37.2	25.3
TOTAL ALL AES	1,133	100.0	27.6	24.7

@ - nonoperative fractures and non-cesarean deliveries only
* - too few observations to present %

Surgical failure (that is, the lack of expected relief of

symptoms or failure of repair) accounted for fewer than 4% of AEs, while nontechnical, or medical complications occurring within two weeks after surgery (acute myocardial infarction, stroke, pulmonary embolus, congestive heart failure, pneumonia, urinary tract infection) were responsible for 7%.

Drug complications were the most common type of AE over all, accounting for 19% of all AEs in the Study. Other nonsurgical AEs that accounted for substantial fractions were procedure-related complications (7%), diagnostic mishaps (8%), and noninvasive therapeutic mishaps (8%). The latter included delays in instituting appropriate therapy, use of outmoded or ineffective therapy, and decubitus ulcers.

These findings are similar to those of the 1977 California Medical Association Medical Insurance Feasibility Study of Potentially Compensable Events (PCE).[28] In that study, Mills et al. found that 67% of PCEs were due to operations and procedures compared to our combined total of 55% of AEs. Other comparable surgical figures from the CMA study were 13% acute technical complications, 10% wound infections, 6% surgical failures, and 6% nontechnical ("medical") complications. Drug complications accounted for 19% in both studies, but the CMA study found only 5% diagnostic errors and 3% of PCEs due to therapeutic management mishaps.

1. Negligence. As described above, of all AEs in New York in 1984, 27.6% were judged to result from negligence. The percentage of AEs due to negligence varied substantially among the different categories, however. Of operative AEs, only 17% were due to negligence, ranging from 13% for wound infections to 36% for surgical failures. On the other hand, 75% of all AEs caused by problems in diagnosis and 77% of non-drug-related,

[28] Mills, D. H., Boyden, J. S., Rubsamen, D. S., and Engle, H. L., Medical Insurance Feasibility Study, Los Angeles, 1977, California Medical Association and California Hospital Association.

noninvasive therapeutic mishaps AEs were judged to be caused by negligence.

2. Disability. In Section III.E we presented population-wide estimates of the level of disability caused by AEs using the ratings of all reviewers (usually two per case). Although this was appropriate for the over all figures, to analyze the individual types of AEs, we assigned a single disability rating to each case. (See Technical Appendix 5.VI.1.)

As noted earlier, the majority of AEs did not result in serious or prolonged disability. The 25% of AEs that caused serious disability (grade 3 or greater) were not distributed evenly among the type of AEs (Table 6.22). For example, nontechnical operative complications (such as pulmonary embolus, heart attack or stroke) were much more likely to result in serious disability (44%) than purely technical complications such as bleeding or wound problems (12%). Late operative complications were associated with serious disability in 36%. Diagnostic mishaps and noninvasive therapeutic mishaps resulted in serious disability in 47% and 35% of AEs, respectively. Drug complications as a group had a low fraction with serious disability (14%). The number of cases was too small to calculate disability for anesthetic complications, nonoperative delivery and fracture AEs, neonatal problems, and falls.

3. Drug Complications. AEs were most likely to result from antibiotics (16%) and antitumor drugs (16%) (Table 6.23). A single class of drugs can cause several different types of AEs. For example, an antibiotic can cause an allergic reaction, damage to the kidneys, or diarrhea. Antitumor drugs frequently interfere with the production of white blood cells and thereby lead to infection.

Table 6.23

Adverse Events By Drug Class

Drug	Number	Percent (Weighted)
Antibiotic	29	16.2
Antitumor agent	31	15.5
Anticoagulant	20	11.2
Cardiovascular agent	13	8.5
Anti-seizure medication	15	8.1
Diabetes drug	8	5.5
Antihypertensive	10	5.0
Analgesics	6	3.5
Asthmatic agent	5	2.8
Sedative, hypnotic	4	2.3
Antidepressant	1	0.9
Antipsychotic	2	0.7
Peptic ulcer medication	1	0.5
Other	33	19.3
TOTALS	178	100.00

Table 6.24 shows the various types of AEs that were caused by drugs. Bone marrow depression was the most common drug complication and was almost always the result of chemotherapy for malignancy. Bleeding, the next most common drug-related AE, was followed by central nervous system complications (seizures, dystonia, confusion). Most bleeding episodes were due to problems in control of anticoagulant medication. Allergic reactions were most commonly cutaneous rashes.

Table 6.24

Types of Drug Complications

Adverse Event	N	Percent (weighted)
Marrow depression	29	16.3
Bleeding	26	14.6
CNS	26	14.6
Allergic/cutaneous	25	14.0
Metabolic	18	10.1
Cardiac	17	9.6
GI	14	7.9
Renal	12	6.7
Respiratory	5	2.8
Miscellaneous	6	3.4
ALL	178	100.0

Negligence was judged to be the cause of the AE in 18% of drug reactions. Because the numbers of individual types of drug-related AEs are too small for stable estimates, we cannot give negligence percentages for each. The types of reactions that were most likely to be judged to be caused by negligence were cardiac (arrhythmias, hypotension), respiratory, and metabolic (mostly hypoglycemia due to insulin misuse). Reviewers seldom considered bone marrow depression to be due to negligence since it is a common and often unavoidable consequence of chemotherapy. Leukopenia is often used as an end point for dosage in the use of certain chemotherapeutic agents. Only 14% of drug complications resulted in serious (>2) disability.

4. Diagnostic mishaps. Three fourths of diagnostic mishaps were judged to be the result of negligence. Ectopic pregnancy was the most frequently missed diagnosis, an error that was frequently judged negligent. Delay in diagnosis of appendicitis and cancer were other common errors. As noted, diagnostic mishaps were associated with substantially higher

fractions of negligence and serious disability than other types of AEs.

5. Complications of procedures. Performance of a procedure was associated with 7% of the AEs. Procedures include nonsurgical invasive therapeutic and diagnostic maneuvers such as cardiac catheterization, esophagoscopy, and insertion of an intravenous catheter. They are generally performed by nonsurgeons and do not require general anesthesia. The fraction of procedure-related AEs due to negligence was similar to that for surgical operations (15%). The most common procedure-related AEs were complications of cardiac catheterization, infections due to insertion of a Foley catheter, and septic complications of intravenous therapy, particularly central venous lines. Serious disability rates in this group were about average.

6. Fractures. The total number of fracture-related AEs (operative and nonoperative reductions) was 54 and constituted 3.6% of the AEs. Major types of AEs were infection (both in the incision and in the bone or joint), prosthesis problems, and healing failures (non-union). The percentage of fracture AEs judged negligent was 21%, but 46% resulted in serious (>2) disability.

7. Postpartum and neonatal injury. Although the unusually high costs of a few settlements of suits related to injuries associated with delivery have led to high malpractice insurance rates for obstetricians, relatively few AEs in our sample were in newborns and their mothers. The total number was only 62 (15 operative deliveries, 18 vaginal deliveries, and 29 neonates), 2.7% of all AEs. This probably reflects the fact that the overwhelming majority of deliveries are uneventful. The percentage of AEs due to negligence in postpartum patients was 26% and in neonatal patients 31%, not noticeably different from

that of the entire sample of 1,133 AEs. The fraction with serious disability was 2% for postpartum patients and 22% for neonates.

B. Types of Adverse Events by Age and Location.

1. Age. We noted earlier that the incidence of AEs increased with age. Patients over the age of 65 had the highest AE rate (5.7%), and although they represented only approximately 27% of the hospitalized population in New York in 1984, they accounted for 43% of all of the AEs. Moreover, 52% of negligent AEs occurred in this age group. We now look at the distribution of various types of AEs in the various age groups. Table 6.25 shows the estimated percentage of each type of AE for each of four age categories.

Table 6.25

Distribution of AEs by Age Group

Age Group	0-15 years		16-44 years		45-64 years		65+ years	
	# of AEs	Percent of AEs in Age Group	# of AEs	Percent of AEs in Age Group	# of AEs	Percent of AEs in Age Group	# of AEs	Percent of AEs in Age Group
Operative Complications								
Wound Infections	14	13.7	66	19.1	53	13.9	27	10.4
Technical Complications	14	13.2	47	14.6	65	17.4	31	9.1
Late Complications	11	7.8	33	9.3	50	10.9	43	11.5
Nontechnical	3	1.6	16	5.1	31	6.3	37	9.3
Failures	4	3.0	33	8.6	12	1.7	9	2.1
Subtotal, Operative AEs	46	39.3	195	56.7	211	50.2	147	42.4
Nonoperative Complications								
Drug Related	15	18.3	40	15.0	70	23.6	53	19.5
Diagnostic Mishap	10	13.2	27	6.9	22	7.0	20	8.6
Therapeutic Mishap	4	2.9	11	3.1	19	5.4	28	12.0
Procedure-related	8	5.9	21	6.1	34	8.8	25	6.5
Falls	.	.	3	0.7	3	0.6	14	5.4
Fractures*	3	1.7	7	1.2	4	0.5	4	1.4
Postpartum*	1	0.4	17	4.6	.	.	.	.
Anesthesia	1	0.5	8	3.5	3	0.6	1	0.2
Neonatal	28	14.6	1	0.1	.	.	.	.
Other	4	3.2	7	2.1	10	3.3	8	4.0
Subtotal, Nonoperative AEs	74	60.7	142	43.3	165	49.8	153	57.6
TOTAL ALL AEs	120	100.0	337	100.0	376	100.0	300	100.0

* - Nonoperative fractures and non-cesarean deliveries only

Drug-related complications were the most common AEs for all age groups except ages 16 to 44, in which they were second to wound infections. After drug complications, patients 65 and over were most likely to suffer AEs from noninvasive therapy and surgical complications.

Surgical failures were a much higher fraction of total AEs in young adults than in the other age groups. In fact, even though patients in the 16 to 44 year age group incurred only 24% of all AEs, they sustained over half of all of the surgical

failures. The operations that most often led to failures recognized in our review were tubal ligation, procedures for back problems, tendon repair, meniscus repair, excision of pilonidal cyst, nasal reconstruction, cervical cerclage, and repair of tibial fractures. These procedures are typically performed in young and middle-aged adults.

Predictably, three types of AEs occurred more frequently with increasing age, or disproportionately in patients 65 and over: nontechnical postoperative complications, treatment (noninvasive) mishaps, and falls. Nontechnical postoperative complications (heart attack, stroke, pneumonia, renal failure, etc.) frequently resulted from impaired organ function, which in turn is more common in older patients. We also found that many of the noninvasive therapeutic AEs resulted from preexisting degenerative disease: congestive failure, arrhythmias, decubiti, and respiratory failure. Finally, elderly patients are typically at increased risk of falling.

2. Location. More AEs (41%) resulted from management in the operating room than anywhere else (Table 6.26). Second were those that occurred in the patient's hospital room (27%). The emergency room, intensive care units, and labor/delivery areas were each the site of approximately 3%. All other locations in the hospital added up to only 5%. Most AEs outside the hospital were attributed to decisions made in the physician's office.[29]

29 Note that the out-of-hospital adverse events measured in our study are only those that resulted in hospitalization.

Table 6.26

Location of Medical Management That Resulted in an Adverse Event

	ADVERSE EVENTS Number of Reviews[a]	Percent of AEs (weighted)	Percent of AEs due to Negligence (Weighted)	Percent of AEs with Disability Severity >2 (weighted)
Location				
Hospital				
Operating Room	1,019	41.0	13.7	22.0
Patient Room	495	26.5	41.1	30.4
Emergency Room	71	2.9	70.4	24.8
Labor and Delivery	123	2.8	27.7	9.8
Intensive Care Unit	53	2.7	30.2	50.4
Radiology	32	2.0	36.9	35.8
Cardiac Catheter	28	0.9	*	*
Ambulatory Care Unit	19	0.8	*	*
Procedure Room	9	0.5	*	*
Recovery Room	7	0.3	*	*
Nursery	10	0.2	*	*
Therapy/Rehabilitation	4	0.1	*	*
Pathology	4	0.1	*	*
Clinical Laboratory	1	0.06	*	*
Hospital Bathroom	1	0.05	*	*
Service Areas	2	0.03	*	*
Other	9	0.3	*	*
SUBTOTALS	1,887	81.2	26.4	25.9
Outside Hospital				
Physician's office	153	7.7	31.2	21.0
Home	48	2.7	11.4	*
Ambulatory Care Unit	32	1.4	53.6	13.7
Nursing Home	11	0.9	*	*
Other	26	1.1	*	*
SUBTOTALS	270	13.8	30.2	17.0
Missing Location Classification	61	5.1	38.6	
TOTALS	2,218	100.0	27.6	24.7

[a] Based on total number of reviews, not number of cases
* Too few observations to present negligence or disability proportions

The fraction of AEs due to negligence was 27.6% for all sites. The negligence rate was low for operating room events (14%) but very high for those in the emergency room (70%). The negligence rate was also high (54%) for AEs in non-hospital ambulatory care units and in the patient's hospital room (41%),

and about average (31%) for those in the doctor's office. In the CMA study, 50% of potentially compensable events (or what we have called AEs) resulted from management in the operating room, but only 8% from management in the patient's room. The frequency in other locations was comparable to ours.

In our study disability also varied widely by site. Not surprisingly, AEs occurring in the intensive care unit had the highest fraction (50%) of serious (>2) disability. Radiology (36%) and the patient's room (30%) also had above average percentages of AEs with serious disability.

C. Physician Error and Negligence

Physician-reviewers were asked to classify errors made in the treatment of patients with AEs. Such a classification, it was hoped, would be a first step in developing methods to reduce the number of AEs. Before examining our results, it may be helpful to the reader to review Table 6.27 in which we present the questions that the physician-reviewers answered.

As can be seen from the table, after determining that an AE was present, reviewers were asked whether the AE was possibly due to a reasonably avoidable error. If the answer was yes, reviewers were asked to classify the error. When reviewers identified more than one type of error, they were asked to rank them. The results of this classification are shown in Table 6.28. The table gives the first choice of each reviewer. Because the table presents the results of judgments by all reviewers, the total number of observations is nearly twice the number of cases in which there was a question of error.

Table 6.27

Evidence for Negligence (AE Analysis Form, page 10)

7.1. Was this AE possibly due to a reasonably avoidable error, or carelessness by either an individual or medical care system, or both?

7.2. CLASSIFY the error. In a few cases more than one error might be present. If that is the case, RANK IN ORDER OF IMPORTANCE (1=most important)

Rank

__ Error in DIAGNOSIS including failure to order or perform appropriate diagnostic procedures or tests

__ Error in PREVENTION of AE due to failure to employ accepted standards of management.

__ Error in PERFORMANCE OF A PROCEDURE OR OPERATION.

__ Error in DRUG TREATMENT

__ SYSTEM ERROR primarily caused by the medical care system.

__ OTHER ERROR not related to above (specify) _____

Table 6.28*

Classification of Error

Classification	Number of Reviews	Percent	Percent Ultimately Judged Negligent
Performance	537	35.2	28.2
Prevention	232	21.9	59.6
Diagnosis	168	13.8	74.7
Drug treatment	87	8.9	52.8
Systems and others	32	2.4	66.0
Unclassified	220	17.9	43.4
Total	1276	100.0	47.3

* Among all AEs with a question of negligence by either reviewer

The term error was not defined in the AE Analysis Form, so

physician-reviewers may have differed in their criteria for assessment. Further, reviewers were asked to identify possible errors in management even when their confidence in the presence of negligence was low because of insufficient information in the medical record. For example, a patient may return to the hospital for a repeat D & C because of retained products of conception following an elective abortion. Although this complication may result from poor technique, it also occasionally occurs after the procedure is done carefully by the best technicians. In the absence of evidence of technical difficulties, the reviewer might assign a low rating of negligence, yet classify it as an error in performance. Or, a wound infection might occur in the absence of any evidence in the record of contamination or breach of sterile technique. By protocol, it would be given a negligence confidence score of 2. (See AEAF Instruction Manual, Appendix 5D, page 20, for explanation.) The wound infection might therefore be classified as an error in performance, but with a low suspicion of negligence.

Neither of these cases would ultimately be classified as negligent AEs because the threshold for a judgment of negligence (average of two reviews > 3.5) would not be reached. Thus, many of the "possibly...reasonably avoidable" errors that were identified by reviewers were not ultimately judged negligent. Indeed reviewers identified one or more possible errors in management in 48% of all AEs, but only 28% of AEs ultimately met our threshold for a judgment of negligence. For each type of error, the percentage of cases that were ultimately judged negligent is also given in Table 6.28.

The most common type of error, accounting for 35% of all errors in this series, was in performance of a procedure or operation. Error in prevention was the next most common, responsible for 22% of errors, followed by diagnostic errors of 14%. Diagnostic errors and prevention errors were most likely to

be judged negligent (75% and 60%, respectively).[30]

From these data we conclude that performance errors are the most common but are the least likely to result from negligence. By contrast, diagnostic errors, while much less common, are very likely to be due to negligence.

Reviewers were also asked to indicate the reasons for the class of error that was identified. (See page 10 of AEAF, Technical Appendix 5.IV.1.) Table 6.29 presents the major reasons that were given by class of error. The number in parentheses is the number of reviewers who identified each type of error, not just the first choice, so these numbers are larger than those in the preceding table. In addition, reviewers were not limited to a single reason so the percentages do not sum to 100.

[30] Because of the small number of cases, the fraction with negligence in System Errors and Others, while high, may not be a representative estimate.

TABLE 6.29

Reasons For Errors

	N	Pct (weighted)
Performance Errors (697)		
Inadequate preparation of patient before procedure	59	9
Technical error	559	76
Inadequate monitoring of patient after procedure	61	10
Use of inappropriate or outmoded form of therapy	24	3
Avoidable delay in treatment	41	7
Physician or other professional practicing outside area of expertise	13	2
Other Performance Error	75	14
Prevention Errors (397)		
Failure to take precaution to prevent accidental injury	178	45
Failure to employ indicated tests	79	23
Failure to act upon results of tests or findings	80	21
Use of inappropriate or outmoded diagnostic tests	6	1
Avoidable delay in treatment	120	31
Physician or other professional practicing outside area of expertise	16	4
Other Prevention Error	77	19
Diagnostic Errors (265)		
Failure to employ indicated tests	134	50
Failure to act upon results of tests or findings	83	32
Use of inappropriate or outmoded diagnostic tests	3	1
Avoidable delay in diagnosis	149	55
Physician or other professional practicing outside area of expertise	17	6
Other Diagnostic Error	24	10
Reason not apparent	16	5
Drug Treatment Errors (153)		
Error in dose or method of use	67	42
Failure to recognize possible antagonistic or complementary drug-drug interactions	10	8
Inadequate follow-up of therapy	65	45
Use of inappropriate drug	38	22
Avoidable delay in treatment	21	14
Physician or other professional practicing outside area of expertise	8	5
Other Drug Treatment Error	18	9
System Errors (68)		
Defective equipment or supplies	8	8
Equipment or supplies not available	8	5
Inadequate monitoring system	8	10
Inadequate reporting or communications	11	26
Inadequate training or supervision of MD or other personnel	15	31
Delay in provision or scheduling of service	10	14
Inadequate staffing	5	6
Inadequate functioning of hospital service	7	8
Other System Error	12	20

Examples of some of the more common errors:

Performance errors were the most common. Many surgical complications fell into this category: injury to organs, retained foreign bodies, failed tubal ligations, etc. Nonoperative technical errors included laceration and bleeding from an artery from insertion of a catheter for cardiac catheterization, pneumothorax secondary to inadvertent puncture of the pleura during placement of a subclavian catheter, etc.

Among prevention errors, failure to take precautions to prevent a complication was the most common reason checked, such as the failure during performance of a gallbladder operation to obtain a cholangiogram (X-ray examination of the bile ducts), an omission that led to subsequent re-operation for stones retained in the common bile duct.

Among diagnostic errors, avoidable delay and failure to employ tests that were indicated were identified as the most common causes. A leading example was failure to diagnose an ectopic pregnancy. Delay in diagnosis could have resulted from failure to obtain an ultrasound examination when the clinical findings suggested the possibility of an ectopic pregnancy, or it could have resulted from failure to perform a laparoscopy when the ultrasound examination was suggestive but not diagnostic.

Drug treatment errors were most commonly caused by inadequate follow-up of therapy. An example was the patient who developed bleeding from anticoagulant medication that was not monitored with sufficient frequency. Errors in dose or method of use were also relatively common.

System errors. Few AEs were classified as system errors. While the low frequency of identification may reflect a low incidence, we suspect that system errors are under-reported. In part this could result from the tendency of physicians not to recognize the interdependence of causative factors and the existence of this type of error. The small number may also

reflect reviewers' difficulty in determining complex interpersonal interactions from data available in the medical record. We suspect that errors classified in other categories would, upon further investigation, be regarded as system errors.

Some of the reasons given for system errors are instructive, however. The most common cause identified was "inadequate training" (e.g., a nurse or house officer who failed to recognize signs of malfunction of a respirator). Next was "inadequate reporting," such as failure to communicate evidence of respirator malfunction to the attending physician once it was recognized. A third system error was "delay in provision or scheduling of a service," such as the failure to receive test results on the blood level of an antibiotic in time to prevent toxicity.

D. Disability as a Function of the Gravity of Negligence

When an AE was judged to result from negligence, we asked reviewers also to make a judgment on the gravity (severity) of the negligence. In making this judgment, reviewers were asked to consider the degree of deviation of treatment from accepted norms, the potential (not actual) consequences of the negligence, the frequency of risk, the degree of emergency, complexity of the case, co-morbidity, and extent to which there is a consensus about correct therapy or diagnosis. A three-point scale was used: 1 - slight, 2 - moderate, and 3 - grave.

Negligent AEs were more likely than non-negligent AEs to result in serious disability. (See Table 6.5 above for definitions of the levels of disability.) We defined **serious** disability as that persisting beyond 6 months (score >2). Only 20% of patients who suffered a non-negligent AE had serious disability, whereas 38% of patients who had a negligent AE had a serious disability. This difference was largely accounted for by

the much higher proportion of serious disability in patients with grade 3 severity of negligence (Table 6.30).

TABLE 6.30

Disability and Gravity of Negligence

Gravity of Negligence@	Number of AEs	Percent of AEs	Percent in Each Disability Severity Class 1	2	3	4	5	6
Non-negligent AEs	853	72.4	64.5	16.0	2.7	5.3	1.8	9.7
Negligent AEs								
Slight (Grades 1-1.5)	76	7.2	77.4	16.2	4.1	0.5	1.9	1.8
Moderate (Grades 2-2.5)	141	13.3	47.0	17.2	4.3	7.8	0.5	21.7
Grave (Grade 3)	63	7.1	19.8	5.9	0.5	7.8	1.1	65.6
ALL	1133	100	59.9	15.4	2.9	5.5	1.6	14.7

@ By case: average of two reviews or the single review when there was only one.

The fraction of AEs resulting in serious disability (>2) increased progressively with increasing gravity (severity) of negligence. Nearly three fourths of patients with grave negligence (Grade 3) suffered serious disability, more than three times the fraction for non-negligent AEs. Likewise, mortality rates increased progressively with increasing gravity of negligence to 66% for AEs with grade 3 negligence, six times the mortality fraction of non-negligent AEs.

While these findings result in part from the fact that potential outcome is a key factor in determining the gravity of negligence, reviewers were asked to separate the results of treatment from disease and to separate their judgments in general from the specific outcome of the individual case. Nonetheless, grave negligence more often than not is associated with severe disability.