Chapter 1 The Hidden Epidemic: The Harvard Medical Practice Study



Malpractice! The word strikes terror in doctors' hearts—and with good reason. All doctors are at risk of being sued when things go wrong, and most doctors are in fact sued at some time in their career, whether or not they did anything wrong. For some high-risk specialties, including neurosurgery, vascular surgery, and cardiology, the percentage sued is very high, and multiple suits are not uncommon. For all doctors, the cost of malpractice insurance is substantial.

So it was not surprising that a sharp rise in medical malpractice insurance premiums in 1985 was viewed by the profession as a "crisis." Such "crises" occurred periodically and were not necessarily associated with either an increase in malpractice claims or in payouts. In this case, the rise had several causes. Because of several years of substantial gains from their investments in the stock market, liability insurance carriers had not raised premiums very much for nearly a decade, but annual payouts (claims settlements) had continued their steady increase.

The need to "catch up," coupled with rising reinsurance rates imposed by overseas reinsurers because of strengthening of the dollar, led companies to raise premiums 40–100% or more. On Long Island, malpractice premiums for obstetricians jumped from \$68,000 per year to \$100,000 [1]. Doctors perceived a crisis.

How big a problem was actual malpractice? No one really knew. No one knew how many people were hurt by negligent care—that is, substandard care. No one knew how many of those patients filed a malpractice suit. Some suspected the number was quite small, but no one knew. Doctors seemed to complain about being sued all the time, but no one knew the facts. No one knew what percentage of malpractice suits were successful. Or how many people suffered from injuries that were caused by medical treatment that was not negligent. No one knew.

And no one had any idea of the costs of medical injury—financial, physical, and emotional: not just the costs of continuing medical treatment, but of lost wages, childcare, home assistance, and long-term disability.

Reflecting on all of this, Howard Hiatt, dean of the Harvard School of Public Health (HSPH), and his good friend, James Vorenberg, dean of the Harvard Law School, conceived of the idea of doing a study to answer these questions. What were the costs of medical injury? How much of it was due to negligence? How successfully did the liability insurance system meet its purported objectives of compensating the injured and deterring bad practice? Did the risk of being sued make doctors more careful and thus reduce the likelihood of patients being harmed? Did the system fairly compensate those who were harmed?

Some experts had expressed interest in no-fault insurance that would pay for all the costs of injury for all patients, irrespective of negligence. Would such a scheme be an economically feasible alternative to litigation? Surely among the faculty of their two schools, they reasoned, there should be enough brainpower to answer these questions and perhaps even develop a better solution.

The place to start, they thought, was with the facts. How many people were harmed by medical treatment in hospitals? What percentage was caused by errors? By negligence? Of those harmed by negligent care, how many sued? What were the costs of medical injury—not just for those harmed by bad care, but for all patients, including those who suffered nonpreventable injuries? How were these costs paid for? All was unknown. All was potentially knowable.

With colleagues, they designed a study to get this information. They used as a model a 1978 study by Don Harper Mills of "potentially compensable events" (PCEs): medical injuries for which a jury might award malpractice money damages. Mills and his team had analyzed 20,684 patient charts of patients discharged from 23 California hospitals in 1974. They found that 4.65% of the patients experienced PCEs of varying severity [2]. Like the California study, the Harvard study would also be a review of medical records. However, Hiatt and Vorenberg believed that to influence policy-makers it needed to be designed as a *populationbased* study, i.e., based on a scientifically designed sample of patients from all types of acute care hospitals serving all patients in a defined geographic area. Only that way would the information be likely to be used for public planning.

Howard's first thought was to seek the approval of the Massachusetts Medical Society, so he approached the president of the society, whom he knew. She thought it was a very bad idea! As we will see with the AMA later, anything that might possibly make doctors look bad was unacceptable. Similarly, Howard found no "takers" among the various private foundations or governmental authorities in Massachusetts.

But, suddenly, there was interest in New York. Howard described the plan to his friend Alfred Gellhorn, who introduced him to the Commissioner of Health in New York State, David Axelrod, whose response was quite positive. Axelrod took him to meet Governor Mario Cuomo, who said, "We've been looking for you! When can you get started?"

Cuomo was struggling with state spending for medical liability claims that was substantial and increasing. Would the Harvard team be willing to do it in New York State? They were delighted to do so— New York's large size and diversity would make the results more credible. When told how much it would cost, Cuomo commented that he expected it to be several times that amount, and he readily authorized an appropriation of \$3.2 million for the study. The Robert Wood Johnson Foundation contributed an additional \$250,000.

Hiatt led the research team. Troy Brennan and Nan Laird led the study design. Troy was just finishing his chief residency in medicine at the Massachusetts General Hospital, but he was uniquely qualified for this study. A Rhodes scholar, he was an honors MD and MPH graduate of Yale Medical School, while simultaneously receiving his JD from Yale Law School. Nan Laird was a professor of statistics, later department chair, at the Harvard School of Public Health, and was a national leader in survey design methodology.

In addition to Brennan, three other physicians were members of this planning group: Benjamin (Bunny) Barnes, a surgeon from Tufts; Howard Frazier, a nephrologist and health services researcher at



Howard Hiatt. (All rights reserved)

HSPH; and Lynn Peterson, a Brigham and Women's Hospital internist. Harvard's William Hsiao (later replaced by Joe Newhouse) and Bill Johnson from Arizona State University were the economists on the team. Paul Weiler, professor at the Harvard Law School, oversaw legal issues. Russell Localio served as project manager, and Ann Lawthers oversaw data management.

The team was about 6 months into the study when, in the spring of 1987, Howard Hiatt approached me to determine my interest in joining them. After 20 years in academic pediatric surgery, I wanted to work in health policy and was finishing a year as a fellow at RAND studying epidemiology, statistics, and health policy in preparation for my new career. At RAND I had become involved in several studies of overuse of healthcare services and was leading a study of underuse. I was returning to Boston and looking for additional opportunities in my new career.

Bunny Barnes, an old friend of mine and surgical colleague from my days as a resident at the MGH and on the staff at Tufts, had recommended me to Howard as someone who could contribute to the study because of my newly acquired analytic skills and substantial clinical experience.

I remember the interview well. In my usual blunt manner, I told Howard that I had no interest in working on malpractice! I had not made a career change and spent a year of my life learning how to do health policy research just to waste it on an issue that was so polarizing and for which I saw no reasonable prospect for change. I was cool to the whole idea.

But Howard explained that the scope of the study was much bigger than malpractice in that it would collect interesting and previously undeveloped data about the substance behind malpractice, medical injury, and also measure its costs to patients. That piqued my interest. I wanted to work on quality improvement; injury and costs were clearly quality issues. At the time, I had not thought much about medical errors. Like most of my colleagues, I considered minor errors unavoidable and serious errors malpractice, the result of incompetence or carelessness. Howard offered me a half-time position, which fit nicely with my commitments to continuing research work at RAND. I accepted his offer, not suspecting it would change my life.

I joined the team just after they had completed the study design. The next major effort was to agree on our definitions, particularly the term for medical injury. Many different terms had been used: "unplanned event," "unanticipated outcome," "unexpected result," "adverse outcome," and, of course, just plain "complication." A common thread was that the injury was beyond the control of the caregivers—and therefore not blameworthy.

Measurement of harm at the time was haphazard, even casual, with little analysis and few records. Even surgical departments, which traditionally had weekly mortality and morbidity ("M&M") conferences, classified deaths from complications as due to errors in judgment, management or technique, or "patient's disease." Remedies recommended were better education for residents and admonishing all to try harder.

This lack of consistent terminology, as well as physicians' concerns about culpability, led to substantial underreporting of iatrogenic injuries. Physicians had few incentives to report. Reporting mechanisms were underdeveloped and largely voluntary. States required hospitals to report deaths but rarely investigated their causes. The Joint Commission asked hospitals to report "sentinel events" (serious injuries), but few hospitals did. Surgical departments had M&M meetings, but neither other departments nor the hospitals kept tabulations or continuing records of iatrogenic injuries. Medical injury was largely invisible, and hospitals and doctors liked it that way. We sought a neutral term that captured all events and to which we could apply a judgment of negligence when indicated. We finally settled on "adverse event." We spent many hours debating its exact definition and ultimately agreed on "an unintended injury that was caused by medical management rather than the patient's underlying disease." The important point was to distinguish harm caused by treatment from harm caused by disease, independent of whether there was an error or negligence. We knew that making this judgment would be difficult for doctors, as it indeed proved to be.

Physicians are very sensitive to any implication that their performance is deficient in any way. Complications were considered either "preventable," which meant someone was to blame, or unpreventable. Most were put in the latter category, which included certain types of complications that everyone knew occasionally happened and were thought to be unavoidable and therefore no one's fault, as well as the occasional unanticipated outcome that seemed to come out of the blue. Our hope was that reviewers could view "adverse event" as a neutral term.

The most common source of injury caused by treatment in the hospital, of course, is a surgical operation, so it was necessary to distinguish this form of planned harm from that due to errors or other failures. Use of the word "unintended" resolved that problem.

We struggled unsuccessfully to devise a reliable way to measure psychological harm, despite its obvious importance, so we restricted our study to physical harm. For "error," we used Reason's definition: "The failure of a planned action to be completed as intended or the use of a wrong plan to achieve an aim." For "negligence," we used the standard accepted legal definition: "Failure to meet the standard of care."

The plan was to obtain data by reviewing medical records of hospitalized patients. We would focus on adverse events that could potentially trigger a malpractice suit. These were injuries that resulted in some degree of disability, temporary or permanent, including death, or were sufficiently severe to prolong the hospital stay. Concurrently, we developed the instruments for data collection and the training materials for record reviewers, both nurses and doctors.

By early 1988, we had settled on our definitions, developed our screening criteria and record review instruments, and constructed

instruction manuals for nurse and physician reviews. We designed a two-step review process: First, registered nurses who were trained in record review for quality assurance would read each randomly selected hospital record in search of one or more of 18 screening criteria (such as post-op fever or transfer to an ICU) that suggested the *possibility* of an adverse event. Second, records that met one or more of the screening criteria would then be independently reviewed by two board-certified physicians to determine if, in fact, there had been an adverse event.

Physicians were asked to rate suspected adverse events on a sixpoint scale based on their confidence—from the information provided in the medical record—that an adverse event had in fact occurred. We used a six-point scale (1 = little or no confidence, 2 = some confidence, 3 = less likely than not, 4 = more likely than not, 5 = highly probable, 6 = virtually certain) to mimic the legal system, which requires a predominance of the evidence with no room for equivocation (50-50).

Reviewers categorized adverse events (AEs) by type (drug reaction, fall, wound infection, etc.) and rated the disability caused by the AE by severity and by duration (temporary or permanent). If an error was found, it was classified as one of five types: diagnostic, prevention, performance, drug treatment, and system. For each type there were additional questions as to the nature of the failure.

Physician reviewers then made a judgment of whether the adverse event constituted negligence, also rated on a six-point scale of confidence. Finally, the AE was rated as to severity (slight, moderate, grave). Except for the well-established definition of negligence, we developed all these definitions and classifications anew, since we found few in the literature and no consensus among physicians or researchers.

The initial screening review of the hospital records was to be performed by a cadre of nurse record reviewers who were skilled at this type of review and were employed by the Hospital Association of New York State (HANYS) which did record reviews as a business. Part of the funding agreement with New York was that HANYS would perform this function for us.

Unfortunately, our project manager had been unable to get agreement on a contract with them, despite many months of negotiations. Time was running out. We were ready to begin the study, but had no one to review the records! Howard turned to me and asked me to see if I could negotiate a contract. I arranged to meet with the head of the HANYS program and flew to Albany on a Saturday morning to meet her over coffee at her home.

Since I had never negotiated a contract in my life, the night before our meeting, I read Roger Fisher's *Getting to Yes*. It was just the ticket. I asked her what they wanted and told her what we wanted, and within an hour we had agreed on the contract and departed friends. At last, the study could begin. We could begin to train these nurses in the use of the record survey instrument.

Finding and training physicians to review the records was more difficult. With help from the NY Department of Health and strong support from the NY State Medical Society, we identified and recruited board-certified internists and surgeons in each of the 51 towns where our study hospitals were. To minimize conflicts of interest, we required that these physicians not be on the staff or have admitting privileges at the hospital whose records were being studied. They were paid the going rate for physician record review.

We met with each group of physicians (typically 4–8 for a hospital) to instruct them in the review process and make sure they understood the definitions. This was a crucial task, since "adverse event" was a new concept for many, and distinguishing treatment-caused injuries from complications of the disease was not something any had ever done.

We also made clear that the term "adverse event" did not mean there had been an error in care. They would find that some were caused by errors and others were not. Part of the purpose of the study was to find out how many there were of each. Despite this caution, we discovered later that many of them considered error the equivalent of negligence, that is, they resulted from the physician not being careful enough. In truth, at that time most of us more or less shared that point of view.

The final design included a random sample of over 31,000 patients who were selected from 51 randomly selected acute care New York hospitals. Government hospitals and mental institutions were excluded. Study hospitals were asked to provide a list of all patients discharged in calendar year 1984. From those lists, patients were randomly selected to reach the appropriate number for each hospital. The hospitals were then asked to make their medical records available for our review.

We were about to launch this enterprise when the leader of data collection, Bunny Barnes, informed Howard that he was leaving the study to go on a round-the-world cruise with his new wife! Howard turned to me to take over. Suddenly, my involvement and time commitment to the study expanded considerably.

We divided the study hospitals into five geographic regions with a similar number of hospitals (10) in each. No one wanted to do the traveling required to supervise data collection in upstate New York, so I volunteered to take it on. From my undergraduate days at Cornell, I knew how beautiful upstate New York was. I looked forward to spending the spring and summer driving around from city to city. By the end of the study, those who chose NYC because it was so easy to get to found that it was rough at times and were envious of my less-stressful experiences.

Data collection began late in the spring of 1988, after training sessions of the physicians at each hospital in each region. We made periodic visits back to oversee the process and personally review a sample of records to make sure they were being reviewed correctly. We later did a formal review of ten charts at each hospital to check reliability of the physician reviews.

Hospitals were very cooperative and retrieved almost all of the records we requested. It is worth noting that at the time we were not required to obtain permission from the patients to review their medical records, something later required by HIPAA rules. This constraint makes it difficult to perform a similar study today.

By mid-1989, we had the results of our initial analysis of the data from the record review in the New York hospitals. In our sample of 30,121 records, we found that 1133 patients had suffered an adverse event, which computed to a serious injury rate of 3.7%, a bit lower than what the Mills study found. Twenty-seven percent of AEs were judged to be due to negligent care. From these data we estimated that in 1984 there were 98,689 adverse events in New York hospitals, of which 13,451 (13.6%) were fatal [3].

There were no differences in rates by sex, but older patients had higher rates. Adverse event rates were substantially higher in some specialties (such as vascular surgery, thoracic surgery, and neurosurgery) than in others. Adverse event rates were higher in large academic medical centers than in community hospitals, but the fraction due to negligence was much lower. Higher negligence rates were found in hospitals with high minority populations.

But the surprising finding was that more than two-thirds of the injuries seemed to be potentially preventable. Reviewers were able to identify specific errors from information in the medical records for 58% of the AEs [4]; subsequent analysis revealed that an additional 11% of AEs resulted from failure to follow accepted practices, raising the total fraction of potentially preventable AEs to 69% [5].

Complications of the use of medications was the most common type of AE, accounting for 19.4% of the total, followed by wound infections (13.6%) and technical complications of surgery (12.9%). Surgical complications accounted for 48% of all adverse events [4] (Table 1.1).

We were very much aware of the limitations of our study—how far it could fall short of our goal of identifying every adverse event and only adverse events. The likelihood is that our numbers underesti*mated* the number of AEs. There were opportunities at each stage for missing an adverse event. At the first step, where nurses identified whether the patient met any of the 18 screening criteria, they undoubtedly overlooked a few. Since our screening criteria were not perfect, some injuries almost certainly occurred that did not trigger one of the criteria. And, since all of our information came from the medical record, if the caregiver chose not to record a symptom or event in the medical record, then we could not measure it. We suspected this was not a small problem, but had no way to quantify it.

At the review stage, physicians also undoubtedly failed to find some AEs that were present. Although some of those would be simply

Table 1.1 Major types of adverse events	Type of AE	% of total
	Medication errors	19.4
	Wound infections	13.6
	Technical complications of	12.9
	surgery	
	Late surgical complications	10.6
	Diagnostic mishaps	8.1
	Therapeutic mishaps	7.5
	Procedure-related	7.0
	Tioeeduie Teluteu	7.0

Adapted from Ref. [4]

overlooked, others likely resulted from inadequate documentation, ambiguous statements, handwriting problems, and the like. These documentation issues were more common in small private hospitals, where records were less standardized and notes were sparse because only the patient's physician writes progress notes. In teaching hospitals, by contrast, there are multiple notes by residents, medical students, and nurses as well.

Bias also probably played a role, leading physician reviewers to under-identify adverse events and over-label negligence. We defined an adverse event as any injury caused by treatment, whether or not there was an error and whether or not it was preventable. This included common and well-accepted complications. While this seems clear on the surface, it was a new concept to our reviewing physicians.

At this time—remember it was 1988—most physicians considered errors blameworthy; they were thought to result from failure to be careful enough and, therefore, negligent. Some physicians had trouble understanding the term "adverse event" as a neutral descriptor, to be applied to all treatment-related injuries, whether or not they were caused by an error.

Thus, despite extensive training and reviews, some physicians still equated "adverse" with error and accordingly might not call an injury an adverse event if there was no error. Some complications were inevitable, the thinking went; they should not be "held against" the physician. Evidence of this kind of thinking is the fact that several types of adverse events that later studies showed to be quite common, such as hospital-acquired infections, falls, and pressure ulcers, were infrequent in our study.

It is unlikely that we were overcounting. Reviewers would not "see" events that hadn't happened! On balance, we believed that our rates, shocking as they were, *underestimated* the true extent of harm. In fact, later studies would bear this out.

The implications of our findings were profound. If our rates were representative, i.e., if adverse event rates in hospitals across the country were similar to what we found in New York State, then nationwide 1.3 million patients were injured by medical care in American acute care hospitals that year, and 180,000 died from these injuries! These numbers were an order of magnitude higher than had ever been suggested. Medical injury was truly a hidden epidemic.

But I was struck with something else: more than two-thirds of the AE were caused by errors and systems failures that we could detect in the medical record. This meant that of the projected 180,000 deaths each year, more than 120,000 were potentially preventable. I was surprised that no one else in the study group found this particularly alarming or of interest. The focus of the study was on malpractice—the costs of injuries and who paid. But it was the fact that two-thirds were potentially preventable that captured my attention. Surely, we should be able to eliminate those—or at least some of them. Preventing these errors and failures could be a huge agenda for improvement. My colleagues disagreed and warned, "Don't go there. The doctors will hate you."

The results of the study were published in two papers in the *New England Journal of Medicine* in February 1991 [3, 4]. It got substantial local coverage in the New York media and some national notice. The New York State Medical Society was not pleased, but made the best of it by claiming that the 1% negligence rate (27% of 3.7% injury rate) was quite low and showed that doctors were performing at a 99% perfect level! [6] But interest in the study faded quickly. No one knew what to do about it, so after a few commentaries from assorted parties, everyone, lay and professional, pretty much quit talking about it.

The Medical Practice Study did one other thing: it determined the feasibility of no-fault insurance as an alternative to the tort system to compensate patients for medical injury. Malpractice suits only compensate patients whose injuries were caused by negligence and who succeed in winning a malpractice suit. Most people don't sue, and most of those who do don't win. The net result is that very few injured patients are compensated by the tort system.

In a no-fault plan, all patients who suffer a treatment-caused injury are compensated for all of its subsequent costs, irrespective of whether the injury was caused by error or negligence. Importantly, these costs also include lost wages, home care, and long-term disability care.

To determine the feasibility of no-fault compensation, we did a follow-up study of the economic consequences of the adverse events. We interviewed the patients from our study who had been injured—or their next of kin if they had died—to determine the long-term effects of the injuries on the victims (such as permanent disability and inability to work), and we estimated their total costs, medical care, lost wages, disability care, etc., over their lifetimes.

From our analysis, we estimated that the total lifetime cost of adverse events in New York State in 1984 was \$3.8 billion in 1989 dollars. Over three-fourths of that cost was paid for by medical insurance or programs such as Medicaid, disability income insurance, and workman's compensation. But the rest was paid by patients.

The cost of a no-fault compensation scheme to compensate for that remainder would be \$878 million per year. In that same year, hospitals and doctors in New York paid \$1.1 billion for malpractice insurance premiums. The obvious conclusion was that we could compensate *everyone* who was seriously injured for *all* of their expenses for less than the amount that doctors and hospitals were already paying for liability insurance that compensated only the small percentage of patients who received a settlement for malpractice [7].

We called for implementation of no-fault insurance. The potential benefits seemed overwhelming. Only 4% of our patients with significant adverse events ever filed a malpractice claim. Multiple studies have shown that fewer than half of malpractice claims ever result in a payment to the patient. Thus, fewer than 2% of the 98,689 patients who were injured in New York in 1984 were likely to receive compensation. By contrast, a no-fault insurance plan would compensate *all* patients who had significant disability.

Our plea fell on receptive ears. David Axelrod, New York's Health Commissioner who commissioned the study, was in full agreement. He got the governor to propose enabling legislation for statewide nofault insurance. It was not to be. Axelrod was tragically disabled by a stroke a few months later, and the state fell onto hard fiscal times. Without his leadership and drive, the legislation perished. An unprecedented opportunity for enlightened government and fairness for victims of medical harm evaporated.

Nonetheless, the Medical Practice Study had a profound impact. Although it was designed to address malpractice, its far greater significance came from the revelation of the horrendous extent of harm that resulted from routine medical care. Here for the first time was indisputable evidence that hundreds of thousands of people were being harmed every year by care intended to help them. And, for the first time, evidence that many of those injuries were potentially preventable. Patient safety was a much greater problem than any of us realized. But it would take some time for this to sink in for the medical profession and its leaders.

References

- 1. Posner J. Trends in malpractice insurance, 1970–85. Law Contemp Probl. 1986;49:37–56.
- Mills DH. Medical insurance feasibility study. A technical study. West J Med. 1978;128:360–5.
- Brennan TA, Leape LL, Laird N, et al. Incidence of adverse events and negligence in hospitalized patients: results from the Harvard medical practice study I. NEJM. 1991;324:370–6.
- 4. Leape LL, Brennan TA, Laird NM, et al. The nature of adverse events in hospitalized patients: results from the Harvard Medical Practice Study II. N Eng J Med. 1991;324:377–84.
- 5. Leape LL, Lawthers AG, Brennan TA, Johnson WG. Preventing medical injury. Qual Rev Bull. 1993;19:144–9.
- 6. Zinman D. Study finds hospitals 'harm' some. Newsday 1990 March 1, 1990;Sect. 17A.
- 7. Johnson WG, Brennan TA, Newhouse JP, et al. The economic consequences of medical injuries: implications for a no-fault insurance plan. JAMA. 1992;267:2487–92.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

