PAP AND CIRCUMSTANCE: WHAT JURY VERDICT STATISTICS CAN TELL US ABOUT JURY BEHAVIOR AND THE TORT SYSTEM

Neil Vidmar

Jury irrationality, the jury lottery, spiraling damage awards, unbridled juror sympathies, the dangerous national sport of punitive damages, the ‘deep pockets’ effect, commonplace verdicts in which fifty or even eighty percent of the award is for pain and suffering, racial and gender biases in verdicts, runaway juries, pro-plaintiff tendencies, capriciousness and unreliability, the defective jury system, exorbitant and unjustified awards, stifled product innovation, raised insurance rates, doctors leaving pediatrics and other areas of practice, nuisance cases, thousands of cases settled for unrealistic amounts because of the threat of even larger jury awards, and an American economy suffering from the litigation costs that must be passed on to consumers.

These are frequently made claims about the behavior and verdicts of juries in personal injury cases, set forth by spokespersons for industry, insurance companies, and professional organizations such as the American Medical Association. Such claims are also disseminated by popular conservative political writers such as Peter Huber and Walter Olson, as

† This article is a revised and expanded version of an article written primarily for social scientists: Neil Vidmar, Making Inferences About Jury Behavior from Jury Verdict Statistics: Cautions About the Lorelei’s Lied, 18 LAW AND HUM. BEHAV. 599 (1994). My colleagues, Randall Bovberg, George Christie, Theodore Eisenberg, Valerie Hans, Richard Lemport, James Levine, Michael Saks, Joseph Sanders, and Clive Seligman made many helpful suggestions as I worked through successive drafts, and so did my research assistants, Jessica Buranosky Lee, Elaine Cohen, and David Landau. I believe that my skeptical approach to the meaning of data is in the spirit of the many contributions Tom Lambert, Jr. has made to the field of law, and I am pleased to dedicate this article in his honor.


2. See generally PETER HUBER, GALILEO'S REVENGE: JUNK SCIENCE IN THE COURTROOM (1991)
well as those writers who lean to the left of the political spectrum. Legal scholars of all stripes similarly make these accusations about personal injury juries. They are repeated in testimony before Congress and by Supreme Court Justices.

In this article, I contend that no justifiable scientific basis exists to support these ubiquitous claims. I attempt to provide an analytic framework for evaluating claims made about juries, particularly those based on statistics about jury verdicts. Application of the framework undermines the claims of those who allege that evidence shows that juries in personal injury cases are biased, unreliable, and incompetent. As I will point out in Section III, verdict data can be useful in answering some questions about jury behavior, particularly when analyzed in conjunction with other variables. Ironically, the data tend to support the view that jury decisions are, on average, temperate and reasonable.

I. THE EVIDENCE OF THE CRITICS

Many contemporary critics of the civil jury system write as if the alleged incompetence and malfeasance of juries is a phenomenon of the last half or even the last quarter of the twentieth century. The charges, however, date at least as far back as the middle of the nineteenth century. In

---

[hereinafter GALILEO’S REVENGE]; PETER HUBER, LIABILITY: THE LEGAL REVOLUTION AND ITS CONSEQUENCES (1983) [hereinafter LIABILITY].


6. See generally Daniels, supra note 1 (discussing testimony before House subcommittee involving alleged increase in jury verdicts); A. Russell Localio, Variations on § 962.258: The Misuse of Data on Medical Malpractice, 13 LAW, MED. & HEALTH CARE 126 (1985) (reporting on jury verdict awards).

1852, in *Haring v. New York & Erie R.R. Co.*, Justice Barculo stated:

We cannot shut our eyes to the fact that in certain controversies between the weak and the strong—between a humble individual and a gigantic corporation, the sympathies of the human mind naturally, honestly and generously, run to the assistance and support of the feeble, and apparently oppressed; and that compassion will sometimes exercise over the deliberations of a jury, an influence which, however honorable to them as philanthropists, is wholly inconsistent with the principles of law and the ends of justice.9

In 1847, an editorial writer in the *Boston Medical and Surgical Journal*10 wrote about conflicting expert testimony that “bewildered” juries,11 and in 1854, a New York doctor asserted that, “[a] single dissenting voice among the surgeons on the stand is enough to turn the scale in favor of the plaintiff toward whom the sympathies of the jury invariably run.”12 In an 1865-66 issue of the *Buffalo Medical Journal and Monthly Review*,13 an editorialist referring to medical malpractice trials said that “sympathy of a jury of citizens is not generally with the doctor, but rather on the side of the poor, ill-advised, unfortunate victim of incurable injury.”14

There are numerous sources of these claims. One source clearly is anecdotes, the selection of cases that are cited as exemplars, that is, typical cases. As other authors have explained, the anecdote cases may not be at all representative of typical cases.15 Indeed, Kenneth DeVille’s study of medical malpractice cases in the nineteenth century strongly suggests that doctors won far more cases than they lost.16 This fact remains true today. In *Medical Malpractice and the American Jury*, I have reviewed a substantial number of studies that demonstrate that across the United States doctors appear to win, on average, two out of three malpractice cases that go to trial.17 One must also remember that the simple fact of high plaintiff-win rates or large damage awards in one or many cases by themselves

11. *Id.* at 54.
12. *Id.*
13. *Id.* at 59.
14. *Id.*
15. *See generally Daniels & Martin, supra note 1, Daniels, supra note 1, at 281; Saks, supra note 1, at 1159.
17. *See generally Medical Malpractice, supra note 1 (contending doctors prevail in two-thirds of medical malpractice cases); Data from Medical Malpractice, supra note 1 (same); Empirical Evidence, supra note 1 (same).*
can tell us nothing about whether verdicts are reasonable or not reasonable. Instead, these figures may properly reflect the fact that plaintiffs deserve to win and that their sustained damages are great.

Another source of the claims are appeals to "intuition" and "common sense." Justice Barcnoe's assertion of the "fact" that juries sympathize with injured plaintiffs is repeated by contemporary sources. In a 1992 videotape produced by the Manhattan Institute for Policy Research, C. Everett Koop, former Surgeon General of the United States, related a hypothetical example involving a child born with cerebral palsy that was not caused by the physician. Dr. Koop stated that, "if they attempt to sue their physician they're very likely to find a sympathetic jury that will award something to the family, not necessarily because they think the doctor is guilty of negligence or malpractice but because their sympathy for the family dictates it." Such intuitive reasoning appeals are pervasive in the writings of Peter Huber and Walter Olson but also appear in the writings of respected legal scholars.

A third, related, source of the claims is the fabricated or misleading anecdote. There are a number of such anecdotes, but probably the one most cited is about the jury that rendered a million-dollar verdict in the case of a woman who allegedly lost her psychic powers after undergoing a CAT scan. This story apparently originated in 1985 in a speech by an insurance company executive. Despite the fact that the ruling was shown to be erroneous, the judge disallowed the woman's claim of loss of psychic powers, and trial evidence showed that, in fact, the plaintiff suffered severe brain damage as a result of a contrast dye that was negligenty administered. Peter Huber in Galileo's Revenge and Walter Olson

18. See Landsman, supra note 9, at 607 (stating juries sympathize with weak plaintiffs).
20. Id.
21. See generally GALILEO'S REVENGE, supra note 2; LIABILITY, supra note 2.
22. See generally WEILER, supra note 5 (noting jurors accustomed to granting huge awards by reaching into deep pockets of malpractice insurers); Sagarman, supra note 5 (discussing jury trial process in cases involving complex scientific evidence). See also CLARK HAIGHURST, HEALTH CARE LAW AND POLICY: READINGS, NOTES AND QUESTIONS 78 (1988) (explaining juries poorly positioned to choose reliably between well argued, but confusing, theories of two sides' experts). In Skidmore v. Baltimore & Ohio R.R., 167 F.2d 54 (2d Cir. 1948), respected federal jurist and scholar, Jerome Frank, argued that "[w]hile the jury can contribute nothing of value so far as the law is concerned, it has infinite capacity for mischief, for twelve men can easily misunderstand more law in a minute than the judge can explain in an hour." Skidmore v. Baltimore & Ohio R.R., 167 F.2d 54, 60 (2d Cir. 1948). Frank cited no evidence to support his position.
23. See Daniels, supra note 1, at 292-93 (discussing anecdotes and horror stories about juries).
24. GALILEO'S REVENGE, supra note 2, at 3-4.
26. The Manufactured Crisis, CONSUMER REP., Aug. 1986, at 546; Fred Strasser, Tort Tales: Old
in *The Litigation Explosion*, writers in *Time, Newsweek,* and elsewhere, and even W. Kip Viscusi's scholarly work, *Reforming Products Liability*, have repeated this discredited anecdote. In another instance, a legal scholar used a judge's comments about "passion and prejudice" regarding a damage award to imply that the jury could not decide the issue of liability, when in fact the trial judge and an appeals court affirmed the jury's liability decision. The lesson is that we should be wary of anecdotes for reasons of accuracy and representativeness.

A fourth source of data about juries was not readily available to jury critics until the mid-1970s; namely, compilations of statistical profiles of jury verdicts. Some of these statistics have come from the closed claim files of liability insurers. Sometimes the statistics are compiled by courts themselves. A third source of statistics comes from verdict reporter publications. Private business enterprises produce these verdict reporters with the primary purpose of providing litigators with information about plaintiff-win rates and the amounts of damages that juries have awarded for specific types of injuries. Often verdict reporters attempt to provide additional information about cases through interviews or questionnaires involving some of the lawyers involved in the case or through the court records. This additional information may include such things as allegations about the nature of the injury, the number of experts involved, and amounts of pre-trial settlement offers.

---

29. See *Daniels*, supra note 1, at 292-93 (discussing jury anecdotes and horror stories).
31. See generally Data from Medical Malpractice, supra note 1.
37. See generally DANZON, supra note 33. See also Samuel R. Gross & Kent D. Syverud, *Getting to No: A Study of Settlement Negotiations and the Selection of Cases for Trial*, 90 MICH. L. REV. 319,
Jury verdict statistics can be divided into two broad categories: representative and non-representative. Non-representative statistics are those compiled on a non-comprehensive or non-random basis. Thus, for example, the reports of Verdict Research, Inc., of Solon, Ohio, are compiled by scientifically unrepresentative searches of newspapers and court records, and by voluntary submissions of journalists or lawyers claiming to be knowledgeable about the cases.\textsuperscript{38} While the number of verdicts upon which the summary statistics are calculated from these sources may involve hundreds or even thousands of cases, they are subject to the same weaknesses of anecdotes. Specifically, they do not allow conclusions about typical verdicts.\textsuperscript{39} For instance, neither journalists nor lawyers are as likely to report verdicts won by defendants or small awards as they are to report mega awards to plaintiffs.

Although non-representative statistics have been cited in the debate about personal injury juries,\textsuperscript{40} they can be dismissed as suffering the same basic weaknesses as anecdotes.\textsuperscript{41} Insofar as their ability to tell us anything meaningful about jury behavior, I dismiss all of the sources of information about juries discussed above as falling within the dictionary meaning of “pap”: that is, “something lacking real value or substance”\textsuperscript{42}

Representative verdict statistics are not as easily dismissed, but as I will demonstrate, they can be very misleading. A number of verdict reporters provide relatively comprehensive data on jury decisions in the jurisdictions that they cover. These data bases, therefore, are not subject to the criticism that they present only a partial and possibly unrepresentative picture of jury trial outcomes. As a consequence, social science researchers at the Rand Corporation’s Institute for Civil Justice,\textsuperscript{43} the American Bar Foundation,\textsuperscript{44} and academic institutions\textsuperscript{45} have conducted studies aimed at de-

\textsuperscript{38} See Daniels & Martin, supra note 36, at 326-28 (questioning reliability and validity of JVR data). See generally Locatio, supra note 6.

\textsuperscript{39} See Saks, supra note 1, at 1242-62 (suggesting insufficient data to draw valid conclusions about legal process).

\textsuperscript{40} See generally Saks, supra note 1.

\textsuperscript{41} Another problem involves research combining highly disparate sources of data without any theoretical rationale, simply to increase sample sizes. Empirical Evidence, supra note 1, at 234.

\textsuperscript{42} The American Heritage Dictionary 898 (2d ed. 1985).


\textsuperscript{45} Clermont & Eisenberg, supra note 34, at 1128-33. See generally Gross & Syverud, supra note 37 (discussing cases selected for trial).
veloping profiles bearing on jury behavior. The findings from some of these studies have been used by those advocating tort reform to support a number of claims. First, over the last quarter-century juries have become more favorable to plaintiffs on the issue of liability and now give awards that are double, triple, quadruple, or even quintuple what they were in the past. Second, juries discriminate against black and female litigants. Third, juries exhibit “deep pocket” tendencies in awarding damages for medical negligence and products liability cases. Fourth, the “pain and suffering” component of jury damage awards is capricious, growing, and constitutes the vast proportion of large tort awards. These various claims about jury outcomes have been ascribed to jurors’ sympathy for plaintiffs and hostility to defendants, to their incompetence, and to their lack of understanding of the economic and societal consequences of their awards.46

The use of jury verdict statistics to attempt to influence public policy can be documented in a number of areas. In 1986 and 1987, two major insurance companies, Higgins & Johnson and Aetna Life & Casualty, ran a series of advertisements in the Wall Street Journal and Time magazine that claimed verdict statistics supported a conclusion that juries were out of control and have helped bring about “the tort crisis.”47 The statistics have been introduced in congressional and state legislative hearings on behalf of attempts to exempt classes of litigation from constitutionally mandated rights to jury trial, to put limits on the “pain and suffering” component of damage awards, and to remove punitive damages from the province of the jury.48 Briefs in Supreme Court cases challenging the appropriateness of punitive damages have also cited to these statistics.49

In the next section of this article, I explain why it is not possible to draw inferences about jury behavior from verdict statistics alone, for the simple reason that very plausible rival explanations of the findings cannot be eliminated. My review of this literature indicates that some, but by no means all, of the social science authors of verdict-statistics studies are aware of some of the methodological problems, but most authors have not

46. See AM. MEDICAL ASS’N SPECIALTY SOC’Y MEDICAL LIAB. PROJECT, A PROPOSED ALTERNATIVE TO THE CIVIL JUSTICE SYSTEM FOR RESOLVING MEDICAL LIABILITY DISPUTES: A FAULT BASED ADMINISTRATIVE SYSTEM 7-8 (1988) [hereinafter AM. MEDICAL ASS’N] (noting that some jury decisions based on insufficient understanding of facts lead to large awards and disadvantage physicians); LIABILITY OF GALILEO’S REVENGE, supra note 2, at 12 (criticizing tort system for finding wealthy defendant equivalent to negligent one). See generally Theodore B. Olson, The Dangerous National Sport of Punitive Damages, WALL ST. J., Oct. 5, 1994, at A17.
47. See generally DANIELS & MARTIN, supra note 1 (discussing jury verdicts’ effect on tort reform); Daniels, supra note 1 (same).
48. See generally DANIELS & MARTIN, supra note 1.
adequately recognized some of the problems. Moreover, some authors have frequently written their reports in ways that minimize methodological cautions and thereby encourage the unsubstantiated conclusions that journalists, legal scholars, judges, and legislators have drawn from them. I will also show that for certain purposes, particularly when used in conjunction with other variables, verdict statistics can provide useful evidence bearing on limited hypotheses about jury behavior. My analytic framework provides the means to distinguish between pap and circumstance. It is organized around problems of methodology.

II. A COMPENDIUM OF PROBLEMS AND THEIR IMPLICATIONS

A. Tip-of-the-Iceberg Problems

Most sources of verdict statistics provide information only about the cases that get to trial. Depending on the type of case, jury trials may account for only between one percent and twelve percent of cases filed in court; the remainder either settle, are dropped by the plaintiff, or are disposed of through judicial or administrative action.50 The iceberg metaphor is useful because it emphasizes the fact that seeing only the “tip” of litigation does not inform us about the shape, or the changing nature, of the larger base upon which the tip rests.51 This raises the specter of threats to what social scientists call “internal validity.”52 Internal validity involves the ability to draw valid inferences about causal relationships between variables. It is jeopardized when there are plausible alternative explanations for the relationship.53

---

50. See Saks, supra note 1, at 1210-12 (stating fewer than 10% of lawsuits require trial for resolution); Neil Vidmar, The Unfair Criticism of Medical Malpractice Juries, 76 JUDICATURE 118, 119 (1992) (stating about 10% of North Carolina malpractice suits filed resulted in jury trials).

51. Indeed, even data and information concerning the cases filed in court cannot tell us the whole story because the profile of filed cases is influenced by rates at which injuries arise, the perception of grievances, the number of disputes that arise out of grievances, and the rates at which disputes are resolved without formal legal action. See William L. Felstiner et al., The Emergence and Transformation of Disputes: Naming, Blaming, Claiming. . . ., 15 LAW & SOC. REV. 631, 649-52 (1981) (discussing importance of studying transformation of disputants’ attitudes); Herbert M. Kritzer et al., The Aftermath of Injury: Cultural Factors in Compensation Seeking in Canada and the United States, 25 LAW & SOC. REV. 499, 536-37 (1991) (concluding cultural differences between United States and Canada account for different propensities to litigate); Saks, supra note 1, at 1198-211 (describing factors impacting reliability of filing data).

52. See generally CAMPBELL & STANLEY, EXPERIMENTAL AND QUASI-EXPERIMENTAL DESIGNS FOR RESEARCH (1963); COOK & CAMPBELL, QUASI-EXPERIMENTATION: DESIGN AND ANALYSIS ISSUES FOR FIELD SETTINGS (1979).

53. See generally CAMPBELL & STANLEY, supra note 52; COOK & CAMPBELL, supra note 52.
1. Changes in Jury "Behavior" Over Time

The Institute for Civil Justice of the Rand Corporation undertook a series of studies that compared jury verdicts over time. As summarized by Mark Peterson, these studies concluded that "plaintiffs are increasingly advantaged in jury trials" because between 1960 and 1985 plaintiff-verdict rates for some types of cases nearly doubled, while inflation-adjusted mean and median awards increased in certain instances by factors exceeding twenty times the amounts of 1960 awards. Many sources have cited these and other verdict-statistics findings as evidence that juries have become increasingly pro-plaintiff and irresponsible.55

Peterson's report notes that "to some degree the trends probably reflect changes in the cases tried to juries," but it also suggests that the trends "could reflect changes in juries' behavior."56 Indeed, either or both explanations are possible and the crucial task is determining whether and to what degree each may be true. Confounding variables in the design of the study, however, make drawing unambiguous conclusions about changes in jury behavior impossible.

Peterson compared verdicts in San Francisco and Cook County, Illinois, at five-year intervals beginning at Time 1 (1960-1964) and culminating at Time 5 (1980-84). The study can be classified as a "time series study" because it measures changes in dependent variables, namely, plaintiff-win rates and amounts of damage awards. As such, it is subject to internal validity problems that can plague time series studies. Two methodological concepts introduced by D.T. Campbell and J.S. Stanley help us understand the problem of drawing the inference that jury behaviors have changed over time. Selection threatens internal validity when there are differences between the types of cases in the comparison groups because we do not know whether the change is due to the variable of interest or to other variables that distinguish the groups. History represents a threat when an observed change in a variable over time might be due to an extraneous event or events that occurs during the time period and thus could be an alternative explanation for the change. Selection may work in conjunc-

54. Peterson, supra note 35, at ix.
55. See generally Am. Medical Ass'n, supra note 46; Physician Payment Review Comm'n, Annual Report (1991); Schwartz, supra note 5; P. Danzon, The "Crisis" in Medical Malpractice: A Comparison of Trends in the United States, Canada, The United Kingdom and Australia, Law, Medicine and Health Care, Law, Med. & Health Care (1990); O'Connell, supra note 5.
56. Peterson, supra note 35, at viii.
57. Peterson, supra note 35, at 6-14.
58. Campbell & Stanley, supra note 52, at 37.
59. Campbell & Stanley, supra note 52, at 37.
60. Campbell & Stanley, supra note 52, at 7.
61. Campbell & Stanley, supra note 52, at 11. "Mortality" or attrition are other terms intro-
tion with history or other confounding factors. To illustrate, I have drawn summary statistics for medical malpractice cases from several places in the Peterson report and combined them to form Table 1.62

**Table 1**

**MEDICAL MALPRACTICE VERDICTS IN SAN FRANCISCO AND COOK COUNTY: 1960-84 (FROM PETERSON, 1987)**

<table>
<thead>
<tr>
<th></th>
<th>Portion Plaintiff Victories*</th>
<th>Median Award in $000$</th>
<th>Mean Award in $000$</th>
<th>No. of Trials$^3$</th>
<th>Percent of All Trials$^4$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>San Francisco Co.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1960-64</td>
<td>.27</td>
<td>$64$</td>
<td>$125$</td>
<td>95</td>
<td>7%</td>
</tr>
<tr>
<td>1965-69</td>
<td>.35</td>
<td>$157$</td>
<td>$306$</td>
<td>88</td>
<td>6%</td>
</tr>
<tr>
<td>1970-74</td>
<td>.43</td>
<td>$124$</td>
<td>$449$</td>
<td>98</td>
<td>7%</td>
</tr>
<tr>
<td>1975-79</td>
<td>.32</td>
<td>$99$</td>
<td>$644$</td>
<td>81</td>
<td>8%</td>
</tr>
<tr>
<td>1980-84</td>
<td>.53</td>
<td>$156$</td>
<td>$1,162$</td>
<td>55</td>
<td>9%</td>
</tr>
<tr>
<td><strong>Cook Co.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1960-64</td>
<td>.25</td>
<td>$35$</td>
<td>$52$</td>
<td>56</td>
<td>2%</td>
</tr>
<tr>
<td>1964-69</td>
<td>.29</td>
<td>$48$</td>
<td>$83$</td>
<td>68</td>
<td>2%</td>
</tr>
<tr>
<td>1970-74</td>
<td>.29</td>
<td>$127$</td>
<td>$605$</td>
<td>100</td>
<td>3%</td>
</tr>
<tr>
<td>1975-79</td>
<td>.33</td>
<td>$141$</td>
<td>$324$</td>
<td>134</td>
<td>5%</td>
</tr>
<tr>
<td>1980-84</td>
<td>.49</td>
<td>$121$</td>
<td>$1,179$</td>
<td>162</td>
<td>5%</td>
</tr>
</tbody>
</table>

*From Peterson (1987) Table 3.1

*From Peterson (1987) Table 3.3

*From Peterson (1987) Table 2.3

*From Peterson (1987) Table 2.4

Note: Median and mean awards were adjusted for inflation and thus the figures involve 1984 dollars.

Compare the malpractice verdicts at Time 1 and Time 5. In San Francisco at Time 1, plaintiffs won twenty-seven percent of jury trials, and at Time 5, they won fifty-three percent, a "win rate" increase of almost one hundred percent. In those cases in which plaintiffs prevailed, the median

62. See generally Peterson, supra note 35, at 11, 13, 17, 21, & 22.
inflation-adjusted award increased from $64,000 to $156,000, the mean, or average, and the mean increased from $125,000 to $1,162,000. (Table 1 shows that very similar verdict trends also occurred in Cook County, but for the moment, ignore Cook County.)

These findings would be compelling if the problems of selection and history did not exist. Consider an elementary example from an imaginary jurisdiction. At Time 1, two types of cases are selected for jury trial: Type A cases are worth $10,000 and Type B cases are worth $100,000. Assume that all defendants have been found liable and that the only issue is the amount of damages. During Time 1, ten Type A and ten Type B cases are tried, and the juries reliably award $10,000 to Type A cases and $100,000 to Type B cases. The mean award for the twenty jury trials is \((10 \times 10,000) + (10 \times 100,000)) ÷ 20 = 55,000\). Assume now that between Time 1 and Time 5, the jurisdiction introduces an alternative dispute resolution program so successful that all Type A cases are settled. Thus, at Time 5 only ten Type B cases go to trial and, as before, juries give each an award of $100,000. The mean award is \((10 \times 10,000) ÷ 10 = 100,000\), an increase of eighty-two percent over the mean award at Time 1. Did jury behavior change? Of course not! Between Time 1 and 5 an outside event occurred (history) and the sample changed (selection). The logic also applies to any changes in plaintiff-win rates.

My hypothetical example has one less obvious implication. One can easily construct scenarios in which absolutely no difference in mean awards occurs between Time 1 and Time 5 but in which juries actually decide cases differently. In the most extreme scenario, consider Time 1 to be the same as described above but suppose at Time 5 all twenty cases go to trial and that juries perversely give the ten Type A cases $100,000 instead of $10,000 and the ten Type B cases only $10,000 instead of $100,000. The mean award would be the same as Time 1, $55,000, and the enormous, perverse change in jury behavior would not be reflected as even a ripple in the averaged awards.

Now, let us apply the lessons from these examples. Go back to Peterson's data in Table 1. Note that at Time 1 San Francisco had 95 malpractice trials, but at Time 5 there were only fifty-five trials, a forty-two percent decrease. Note, also, that despite this attrition in absolute numbers, malpractice trials actually constituted a greater proportion of the trial docket, increasing from seven percent to nine percent. Clearly, unusual changes occurred in the selection of cases that went to trial in San Francisco between 1960 and 1985. The immediate lesson is that selection and history are major confounding problems in any attempt to draw conclusions about changes in jury behavior.63 The “disappearing” cases may

---

63. Other evidence indicates that the number of malpractice lawsuits in San Francisco increased
have produced the changes in awards and verdicts rather than anything the juries did. To say, as Peterson does, that the data may reflect either changes in the cases tried or changes in jury behavior does not mean that each is equally probable nor does it allow us to conclude that both changes have probably occurred. Scientifically, the verdict data do not allow us to draw any conclusions. Without controlling for outside events or changes in the cases that worked their way to the tip of the litigation iceberg, it is not possible to draw conclusions about trends or lack of trends in jury behavior. In other words, absent additional data, we simply cannot determine whether juries are deciding cases differently or whether they are deciding different cases.

2. Differences Between Jurisdictions

Other authors have attempted to compare jury behavior across jurisdictions. Once again the Rand Corporation studies help to illustrate the problem arising from the fact that verdicts represent only the tip of the iceberg. The data in Table I from Cook County illustrate that the trends in verdicts from Time 1 to Time 5 were essentially the same for malpractice cases in Cook County as they were in San Francisco. Specifically, from 1960 to 1985, plaintiff-win rates almost doubled and median and mean awards were larger by factors of 3.5 and 22.7, respectively. When we look at the number of cases going to trial in Cook County, however, the pattern is exactly the opposite of that in San Francisco. While San Francisco’s malpractice trials decreased by almost half, Cook County’s cases increased by almost two hundred percent, and their percentage of the trial docket increased from two percent to five percent. This fact, of course, raises the same history and selection problems for drawing conclusions about changes in jury behavior in Cook County between Time 1 and Time 5 as did the data for San Francisco. Primarily, different numbers of cases are being compared and they may well be different types of cases or different proportions of case types. The data, however, also raise the question of whether the two counties were equivalent in the cases chosen for trial at Time 1 or at any of the other time periods, another classic selection confound. In short, even though the “trends” in verdicts are similar between San Francisco and Cook Counties, some very different things were going

---

over the period and that a vigorous alternative dispute resolution program removed many cases from the trial docket. Peterson, supra note 35, at vi-ix. Over time, many things may be in flux and thereby affect the kinds of cases that go to trial, for example, new ADR programs, court docket backlogs, changes in local legal culture and in the judiciary, new court rules, and changes in substantive laws.

64. See generally Peterson, supra note 35. See Daniels, supra note 1, at 299-304 (discussing Jury Verdict Research, Inc. national poll); Daniels & Martin, supra note 36, at 328-47 (discussing evidence from national polls).
on in those courts that affected the cases selected for trial, and without additional evidence, we cannot legitimately ascribe the changes in either win rates or average awards to changes in jury behavior.

Daniels' research comparing jury verdicts in medical malpractice cases in forty-two jurisdictions in eleven states shows that median jury awards are not consistently enormous in all jurisdictions, as some critics of malpractice juries have claimed.65 As the second implication of my hypothetical example illustrates, however, we cannot conclude that juries behave the same or differently in these forty-two jurisdictions, because the trial cases may have been selected differently in different jurisdictions. These juries may just have been deciding different cases.

3. Differences Across Cases Types: The Confounding Effects of Selection on "Deep Pockets" Evidence

A number of studies have found that awards in medical malpractice and products liability cases were substantially larger than awards in automobile negligence cases.66 Many authors have interpreted these findings as evidence of a "deep pockets" effect in which juries assess larger awards against wealthy defendants because they can afford to pay.67 The "deep pockets" conclusion is plagued by the "apples and oranges" problems that I will discuss in Section B, but the problem of selection also exists. For example, because the case law is so settled and insurance companies routinely process claims, only between 1 and 2% of automobile injury cases go to trial; in contrast, between 7 and 12% of medical negligence cases go to trial.68 Because verdict data give no information on the baseline of suits from which trials emerge, many characteristics may distinguish trials across categories of case types. The fact of different trial rates raises the plausible alternative hypothesis that selection jeopardizes attempts to compare verdicts in auto negligence with malpractice or products liability cases. Differences in trial outcomes cannot be automatically ascribed to biases in jury behavior. Once again, juries may be deciding very different types of cases.

65. See Daniels, supra note 36, at 339-43 (asserting size of medical malpractice awards gives them visibility beyond their numbers).
66. See generally Chin & Peterson, supra note 35; Peterson, supra note 35; Randall R. Bovbjerg et al., Juries and Justice: Are Malpractice and Other Personal Injuries Created Equal?, 54 LAW & CONTEMP. PROBS. 5 (1991); Danzon, supra note 55; Hammitt et al., supra note 43.
68. See generally Gross & Syverud, supra note 37; Kritzer et al., supra note 51; Saks, supra note 1.
4. Differences Among Juries and Judges

A study by Kevin Clermont and Theodore Eisenberg, using federal court verdict statistics, showed that judges found for plaintiffs more frequently than juries.\(^\text{69}\) Clermont and Eisenberg indicated that most of the difference was probably because of the types of cases selected for jury or bench trials.\(^\text{70}\) They also speculated, however, that a portion of the variance was probably due to juries being more conservative on the matter of liability.\(^\text{71}\) Samuel Gross' commentary on the study ascribed almost all of the variance to case selection,\(^\text{72}\) but a report of the study in the Wall Street Journal placed emphasis on the interpretation that juries are more conservative in their behavior.\(^\text{73}\) Without knowing about the way parties select cases for trial, however, there exist no grounds for concluding that the percentage of the variance caused by jury conservatism is zero, or some other figure. Indeed, my hypothetical example of type A and B cases suggests that because of selection differences in which cases go to trial, it is theoretically possible that juries are more pro-plaintiff than judges.\(^\text{74}\) The summary statistics that fail to inform us about the base of the litigation iceberg obscure this possibility.

5. Recent Developments in Acknowledging the Iceberg Problem

The tip-of-the-iceberg problems are increasingly recognized as a serious limitation of verdict-statistics studies. Recent articles have discussed the problem in various ways.\(^\text{75}\) Verdict statistics studies, however, are often

\(^{69}\) See Clermont & Eisenberg, supra note 34, at 1133-35 (noting categories in which plaintiffs have greater success before judges).

\(^{70}\) See Clermont & Eisenberg, supra note 34, at 1128-33 (discussing selection effect theory).

\(^{71}\) See Clermont & Eisenberg, supra note 34, at 1162-66 (noting distinctions between judges and juries on liability and damages).


\(^{73}\) See Milo Geyelin, Judges, More Than Juries, Side with Personal Injury Plaintiffs, WALL ST. J., July 23, 1992, at B1, B7 (commenting that judges more prone to side with personal injury plaintiffs).

\(^{74}\) For instance, if the cases that are selected to go before juries, as compared to those selected to go before judges, contain significantly more cases that should be won by defendants, then the lower plaintiff win rates before juries could still be reflecting a bias in favor of plaintiffs. As an example, consider medical malpractice cases. Clermont and Eisenberg found that plaintiffs won 52 percent of the cases tried before judges but only 30 percent of cases tried before juries; a differential of 22 percent. See Clermont & Eisenberg, supra note 34, at 1175 (charting win rates in judge and jury trials by category). Suppose, however, that we had some way of determining the "correct" outcome according to a reliable, neutral criterion. Suppose that the plaintiffs should have won 62 percent of the judge trials but only 20 percent of the jury trials. The 52 percent win rate would therefore indicate that the judges favored defendants despite the fact that the juries favored plaintiffs even though plaintiffs won 22 percent more often when judges decided their cases.

\(^{75}\) See Bovbjerg et al., supra note 66, at 9 (noting selection power of plaintiffs' attorneys);
beset by other methodological problems that have received less attention.

B. The Apples-and-Oranges Problem

Assume for analytical purposes that no tip-of-the-iceberg problems exist and that, for example, a random and equal percentage of two categories of lawsuits are chosen for trial, such as medical malpractice and automobile negligence. Confounding variables that raise very plausible alternative explanations still plague attempts to make verdict comparisons of the two types of cases and to ascribe differences in those verdicts to differences in the way that juries treat plaintiffs or defendants. The two types of cases differ on multiple dimensions, and these other dimensions, singly or in combination, may result in the observed differences in verdicts. Rather than jury biases, the verdict differences may result from the fact that the juries are hearing very different evidence or responding to different substantive law. Comparisons are being made between apples and oranges.

1. Race and Gender Discrimination

One of the Rand Corporation studies of Cook County found that black plaintiffs and defendants lost their cases more often than did whites and that black plaintiffs received lower awards than whites.76 Scholarly articles77 and the mass media78 have cited these findings as evidence that juries discriminate against black defendants.

In the initial report of the findings, Audrey Chin and Mark Peterson conceded that unaccounted-for variables, particularly the way cases are selected, could potentially explain the differences.79 In a later report, however, Peterson ignored this qualification and stated that “studies also found that juries were not always even-handed: Blacks in Cook County and women in San Francisco did not fare as well with juries. . . .”80 Chin

---

Daniels & Martin, supra note 44, at 70 (noting problem of basing policies on anecdotal evidence); Gross, supra note 72, at 1180 (observing fault in verdict statistic study); Saks, supra note 1, at 1189 (noting difference regarding “tip of injury iceberg”); see also Deborah K. Hensler, Reading the Tort Litigation Tea Leaves: What’s Going on in the Civil Liability System?, 16 JUSt. Sys. J. 139, 150 (1993) (finding insufficient data to support analysis of trends in jury verdicts).

76. See generally CHIN & PETERSON, supra note 35.
77. See Alschuler, supra note 5, at 1827 (noting wide disparity).
78. See Janice C. Sampson, White Justice, Black Defendants: Many Believe the Courts Still Biased, TIME, Aug. 8, 1988, at 17 (noting civil suit awards vary by two-thirds or one-half depending on race).
79. See generally CHIN & PETERSON, supra note 35.
80. PETERSON, supra note 35, at v. In fairness, Peterson’s report notes and discusses possible alternative explanations for the data, including changes in case selection. PETERSON, supra note 35, at viii-ix. The use of the phrase “not always even handed,” however, clearly implies bias in the jury’s behavior. PETERSON, supra note 35, at viii-ix.
and Peterson should have heeded their original qualifications, because
differential selection of cases for trial constitutes a plausible alternative ex-
planation. We know nothing about the rates at which black and white
plaintiffs brought suits or settled them and, once again, verdict data tell us
very little about the problem.81

There is another problem in addition to the selection confound, howev-
er, namely that the verdict reporters from which the data were compiled
do not provide enough information about the cases to tell us if Peterson
actually compared apples and oranges. There exist plausible alternative
hypotheses that could explain the differences in verdict outcomes. While
these hypotheses may involve social or institutional factors related to race,
they provide plausible alternative explanations contrary to the hypothesis
that juries discriminate against blacks. For instance, blacks may retain less
competent lawyers, blacks may not be able to afford the economic or
other experts to demonstrate liability or the costs of their injury, and
blacks may have lower incomes and therefore cannot demonstrate econom-
ic losses as large as those of whites.82 This juries may be responding to
different evidence and the quality of the evidence and arguments rather
than engaging in discrimination.

Similar plausible alternative hypotheses exist to explain Peterson’s find-
ing that female plaintiffs received lower awards than males.83 One cannot
scientifically infer that the cause of lower awards results from discrimina-
tory behavior on the part of juries because it is equally plausible to hy-
pothesize that women had lower incomes and, therefore, were not as able
to show economic loss; that they could not afford experts; and that they
could afford to hire only less competent lawyers; or that because of eco-
nomic necessity, more of them settle at discounted rates rather than wait
for trial. Jury verdict statistics provide little or no information about such
confounding variables, but there is a very reasonable probability that these
confounds exist.

81. The Rand data do indicate that smaller numbers of black plaintiffs were represented in trial
cases. See generally CHIN & PETERSON, supra note 35.

82. Professor Michael Saks, in a personal communication, has brought to my attention another
possibility. As a result of black’s lesser wealth, they may have greater needs for immediate money,
and thus plaintiffs settle for less money rather than face trial. Consequently, larger stakes cases involv-
ing blacks may be removed from the trial docket, thereby making their average damages at trial, on
average, smaller. This hypothesis would qualify as a selection confound.

83. See PETERSON, supra note 35, at v (noting discrepancy in verdict awards for women in San
Francisco).
2. Juries Versus Judges

The Clermont and Eisenberg study\(^8^4\) comparing jury versus judge verdicts is relevant because, aside from the selection confound already discussed, there exists evidence that lawyers try cases differently depending on whether the factfinder is a jury or a judge. Gordon Bermant and his colleagues conducted interviews with attorneys in federal court.\(^8^5\) These attorneys indicated that in bench trials they presented every relevant detail of the case, whereas in jury trials they culled the evidence, abbreviated it, and presented it in a different manner. Furthermore, judges have access to information that may be kept from the jury. Thus, when verdict data show that judges and juries differ on verdicts, it does not necessarily mean that one or the other made mistakes or weighed evidence differently. They may simply have been responding to different evidence.

3. Automobile Negligence Versus Medical Malpractice, Products Liability, and Other "Deep Pockets" Cases

As described earlier, automobile negligence cases have frequently served as a standard against which verdicts in medical malpractice and products liability cases are judged.\(^8^6\) The Rand Corporation studies indicated that even when juries considered injury severity, medical malpractice plaintiffs and products liability plaintiffs received awards several times greater than those received by automobile injury plaintiffs.\(^8^7\) Randall Bovbjerg and his colleagues found that awards in medical malpractice cases were approximately six times larger than those in automobile negligence cases.\(^8^8\) When these researchers controlled for injury severity, amount of reported economic damages, and other factors, malpractice awards remained approximately three times larger than those in the automobile cases. While the authors of the above studies have taken some cognizance of the tip-of-the-iceberg problems discussed earlier, they have given much less attention to the other variables that distinguish the case types.

In another article I have delineated a partial listing of some of the differences between medical malpractice and auto injury cases.\(^8^9\)

---

84. See generally Clermont & Eisenberg, supra note 34.
86. See generally AM. MEDICAL ASS'N, supra note 46; Peterson, supra note 35; Physician Payment Review Comm'n, supra note 55; Weiler, supra note 5; Kird B. Johnson et al., A Fault-Based Administrative Alternative for Resolving Medical Malpractice Claims, 42 VAND. L. REV. 1365 (1989).
87. See generally Chin & Peterson, supra note 35. See also Hammitt et al., supra note 43, at 752 (noting disparity in verdicts).
88. See generally Bovbjerg et al., supra note 66.
89. See generally Medical Malpractice, supra note 1; Data from Medical Malpractice, supra...
bile cases often involve multiple plaintiffs, the driver and passengers of the second car, but usually a single defendant, the allegedly negligent driver of the first car. Malpractice cases, however, typically involve a single plaintiff, the injured patient, and multiple defendants (such as doctors working as a team to provide different aspects of the health care). Automobile cases often involve strangers, whereas malpractice cases involve a professional relationship between patient and physician. Automobile cases usually involve a single theory of liability, whereas malpractice cases frequently involve multiple theories. Automobile cases may involve contributory negligence, which is not claimed as often in malpractice cases. Medical malpractice lawyers tend to be specialists who carefully screen cases and invest heavily in experts, whereas generalist lawyers who often call few or no experts litigate automobile cases. Automobile cases frequently concern disputes about damages, whereas malpractice defendants mostly dispute liability. Finally, rather than the "reasonable man" standard used in automobile and most other negligence cases, malpractice juries determine negligence according to whether the physician violated accepted standards of practice.90 The crux of the matter is that juries hear very different cases in medical and automobile negligence trials and decide them under different legal standards.

Products liability trials differ from automobile and medical malpractice cases on other dimensions.91 Few of these confounding variables that distinguish cases are contained in even the more comprehensive verdict-statistics sources and, therefore, cannot be taken into account in statistical analyses even if authors took notice of them.

4. Within Case-Type Variability of Damages:
The Jury Unreliability Claim

The insight that many dimensions of trials differ across case types and result in juries hearing very different trials also cautions against drawing conclusions about jury variability within case types. Some studies have concluded, for example, that while the size of awards correlates positively with the severity of injury, within categories of severity there remains great variance in damages for products liability,92 medical negligence,93 and other tort cases.94 Some have interpreted this finding as showing the

note 1.
91. WEILER, supra note 5, at 19-26. See generally Daniels & Martin, supra note 44.
92. See generally VISCUSI, supra note 30.
94. Randall Bovbjerg et al., Valuing Life and Limb in Tort: Scheduling "Pain and Suffering." 83
unreliability of jury decision-making. Some variability undoubtedly accrues, however, from variability in the trial evidence itself. As any observer of the legal system will attest, no two trials are alike. Indeed, even a retrial of the same case with the same lawyers may produce a very different trial. Some of the most salient dimensions are the skill and experience of the attorneys, the testimony of civilian and expert witnesses, idiosyncracies of judges and their rulings, documentary evidence, and demonstrable economic losses. Although Shari Diamond has shown that the reliability of juries is probably comparable or superior to decision-making in other spheres of human activity, there still exists some unreliability in jury decision-making. The error in drawing conclusions about jury unreliability solely from verdict data, however, is that it ignores the other factors that could plausibly account for verdict variability.

In fact, a very important recent study by Professor Frank Sloan and his colleagues lends strong support to this alternative interpretation. They obtained a sample of Florida medical malpractice cases involving claims arising out of birth injuries and emergency room injuries. Utilizing medical records, liability insurer reports, and extensive interviews with claimants, Sloan and his colleagues made careful estimates of both past and future costs of health care, past and future income losses, family members’ loss of income resulting from the need to care for the injured person, alterations to homes, special transport vehicles, special schools, and other economic losses. Sloan based the economic value of all losses on government and other standard sources; the calculations took life expectancy into account and made deductions for government benefits and services. The estimates included no figures for non-economic losses due to pain and suffering or loss of enjoyment of life. Even when the severity of the injury sustained was taken into account, there was substantial variability of economic losses between individual plaintiffs. For example, among the most seriously injured children about forty percent had economic losses estimated in the range of $1.2 to $1.4 million but about ten percent had

---

95. Id. See generally AM. MEDICAL ASS’N, supra note 46; WEILER, supra note 5; Peter H. Schuck, Mapping the Debate on Jury Reform, in VERDICT: ASSESSING THE CIVIL JURY SYSTEM (Robert E. Litan ed., 1993).


99. See id. (outlining calculating of economic losses involved in representative medical malpractice actions).
losses in the range of $2.4 million to more than $3 million.\textsuperscript{100} Excluding the least seriously injured persons, about twenty-eight percent of claimants in emergency room cases incurred losses ranging from zero to $200,000, but about fourteen percent incurred estimated losses exceeding $3 million.\textsuperscript{101} For patients who died, estimated economic losses ranged from near zero to about $1.2 million. As striking as these figures are, they do not include variability that may occur at trial due to the skill of lawyers, the quality and quantum of experts, or other factors.\textsuperscript{102} Rather than jury caprice or incompetence, variability in the actual economic evidence at trial—apples in one case, oranges in another—represents a plausible explanation of variability in damage awards.

5. \textit{Summary}

It is not scientifically legitimate to ascribe variability in verdicts solely to bias or caprice in jury behavior when juries may be legitimately responding to a host of other variables that distinguish the trials that they must decide. Verdict statistics data sources do not document these other factors and, therefore, the statistical analyses do not account for the confounding variability cause by these factors.\textsuperscript{103} The apples and oranges problem does not allow us to conclude that juries are free from bias or caprice, but it also does not allow us to conclude the contrary.

C. \textit{Unreliable Subtrahend Problems}

The "pain and suffering" component of jury awards represents one of the most criticized aspects of jury behavior. With reference to medical malpractice cases, Paul Weiler has written that "the most troublesome feature of large tort verdicts is the amount of damages awarded for pain and suffering, not for direct medical costs."\textsuperscript{104} In 1988, the American Medical Association claimed that "for larger claims . . . non-economic injury is often the largest component of the award."\textsuperscript{105} Likewise, in

\begin{itemize}
\item \textsuperscript{100} See \textit{id.} at 147 (charting distribution of economic loss by group).
\item \textsuperscript{101} See \textit{id.} at 143 (providing chart that outlines losses in emergency room cases).
\item \textsuperscript{102} See generally NEIL VIDMAR, THE AMERICAN JURY AND MEDICAL MALPRACTICE (manuscript on file with the author).
\item \textsuperscript{103} Through a statistical technique called regression analysis, the effects of confounding variables can be partitioned out to assess the relationships between the remaining variables. See Bovbjerg et al., supra note 94 (illustrating use of regression analysis techniques). Where possible confounding variables exist, however, without a data base to identify or measure them, conclusions about the relationships between the remaining variables lacks validity.
\item \textsuperscript{104} Weiler, supra note 5, at 54. According to Weiler, claiming approximately 50 percent of total awards is for pain and suffering. Weiler, supra note 5, at 55.
\item \textsuperscript{105} See generally AM. MEDICAL ASS'N, supra note 46 (claiming that up to 80 percent of total awards for pain and suffering).
\end{itemize}
1992, the Physician Payment Review Commission asserted that "[m]uch of the unpredictability and inconsistency of malpractice awards is due to noneconomic damages (i.e. pain and suffering), which constitute about half of total payments." Patricia Danzon has claimed that between the period 1959-1979 and the period 1980-1984, malpractice plaintiffs' "share of total payment for pain and suffering increased more than seven-fold" and that, "payments for pain and suffering in excess of $100,000 for medical malpractice cases are increasing disproportionately, relative to payments for economic loss." W. Kip Viscusi has stated that, with respect to products liability cases, "pain and suffering amounts are often the largest part of the damages award." Similar claims have been made with respect to tort cases generally and have been used to bolster arguments for legislative or judicial limits on the amounts that juries can give for non-economic damages, an enterprise that appears to have had considerable success in that a large number of jurisdictions have placed caps on "pain and suffering."

The foundation behind these claims derives in large part from studies that have used verdict data to estimate the percentage of jury awards that can be ascribed to "pain and suffering," as opposed to economic damages. The studies generally conclude that pain and suffering awards positively relate to the seriousness of injury, but that they constitute a large percentage of the total award and show great variability within categories of injury seriousness.

Whatever the empirical substantiation of the various arguments about pain and suffering might ultimately prove to be, the studies that purport to show excessive pain and suffering awards have serious methodological problems that confound interpretive efforts. Consider how the estimates for the studies were derived. The researchers used verdict reporters or insurers' closed-claim files that report the total jury verdict and also provide some information about economic damages. Estimates of non-economic damages were derived by subtracting the economic damages (the "subtrahend") from the total awards (the "minuend"), to yield a "remain-

106. See generally PHYSICIAN PAYMENT REVIEW COMM’N, supra note 55.
107. Danzon, supra note 55, at 49.
109. See generally GALILEO’S REVENGE, supra note 2; Bovbjerg et al., supra note 94; Priest, supra note 67; George L. Priest, The Role of the Civil Jury in a System of Private Litigation, 1990 U. Chi. LEGAL F. 161; Schwartz, supra note 5.
110. See generally Bovbjerg et al., supra note 94; Schwartz, supra note 5.
111. See generally M. PETERSON, COMPENSATION OF INJURIES: CIVIL JURY VERDICT IN COOK COUNTY (1984); P. Danzon, Report on Awards for Noneconomic Losses, in MEDICAL MALPRACTICE POLICY GUIDEBOOK (Henry Manne ed., 1985); Bovbjerg et al., supra note 94; Danzon, supra note 55.
112. See generally Peterson, supra note 111; Danzon, supra note 111, Bovbjerg et al., supra note 94; Danzon, supra note 55.
der," usually labelled simply as "pain and suffering" by the researchers. While this procedure may appear reasonable, it obscures three very serious problems, bearing on the subtrahend that I will label as "missing data," "the dim bright line," and "the contentious criterion."

1. Missing Data

Even with verdict sources that claim to be relatively comprehensive with respect to the trials covered, data on the economic costs are often missing to an extensive degree. In her study of insurers’ closed-claims files, Danzon stated that the data “only report the insurance company’s estimate of economic loss” and “the data on earnings were particularly poor.” Viscusi conceded similar problems with his closed-claim data on products liability cases:

The loss figures . . . are reported losses, not actual losses. The amount of the loss is recorded in the data set by the insurance company. To the extent that this variable reflects the insurance company’s loss estimates, the financial loss will be understated . . . [and] bias may also vary with loss level. Very large loss claims with substantial long-term medical costs may have a very wide possible variance, so that the propensity for underestimation by the insurance company and overestimation by the claimant will be especially great.

Peterson acknowledged that “occasionally the [verdict statistics] Reporter includes a plaintiff’s income and the length of time that his or her income was lost because of an injury.” Presumably, information about future medical expenses or lost income were absent in Peterson’s data because it was not discussed in the report of study. Frank Sloan and Chee Ruey Hsieh noted similar problems with missing data. My own experience with the closed-claim files of three separate North Carolina medical liability insurers is entirely consistent with these conclusions. Sometimes the economic loss data are presented only in the most general terms, and often they are missing altogether.

Bovbjerg and his colleague’s study clearly demonstrate the difficulties


114. DANZON, supra note 33, at 40.

115. VISCUSI, supra note 30, at 102; Viscusi, supra note 113, at 206.

116. PETERTERSON, supra note 111, at 4.

117. See generally Sloan & Hsieh, supra note 93.

118. See generally MEDICAL MALPRACTICE, supra note 1; Vidmar, supra note 50.
arising from missing data. Those authors reported data on the means, medians, variances, and ranges of total verdicts as a function of the alleged seriousness of the injury. Then, in a second table, they reported summary statistics only for non-economic damages (total verdict minus economic damages). Working from those tables, I have compared the median awards in my own table, along with my calculations about the percentage of total award represented by "pain and suffering."

### Table 2

**JURY VALUATIONS OF TOTAL AND NON-ECONOMIC LOSS IN PERSONAL INJURY CASES: DISTRIBUTION BY SEVERITY OF INJURY (IN THOUSANDS OF 1987 DOLLARS)**

<table>
<thead>
<tr>
<th>Severity Description</th>
<th>Median total award $</th>
<th>Median non-economic award $</th>
<th>Non-economic award as percentage of total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Emotional: e.g., fright, no physical damage</td>
<td>$37</td>
<td>$25</td>
<td>68%</td>
</tr>
<tr>
<td>2. Temp. insig&quot; e.g., lacerations, minor scars</td>
<td>$7</td>
<td>$2</td>
<td>29%</td>
</tr>
<tr>
<td>3. Temporary minor: e.g., infections, misset fractures; fall, recovery delayed</td>
<td>$26</td>
<td>$9</td>
<td>35%</td>
</tr>
<tr>
<td>4. Temporary major: e.g. burns, left surgical material; drug side effect; brain damage</td>
<td>$82</td>
<td>$36</td>
<td>44%</td>
</tr>
<tr>
<td>5. Permanent minor: e.g., loss fingers, damage to organs; non-disabling injuries</td>
<td>$106</td>
<td>$46</td>
<td>43%</td>
</tr>
<tr>
<td>6. Permanent significant: e.g. deafness, loss of limb, eye, one kidney, or lung.</td>
<td>$438</td>
<td>$292</td>
<td>67%</td>
</tr>
<tr>
<td>7. Permanent major: e.g., paraplegia, blindness, loss two limbs, brain damage</td>
<td>$1,422</td>
<td>$1,642</td>
<td>115%</td>
</tr>
<tr>
<td>8. Permanent grave: e.g. quadriplegia, severe brain damage, lifelong care; fatal prognosis</td>
<td>$4,214</td>
<td>$1,832</td>
<td>44%</td>
</tr>
<tr>
<td>9. Death</td>
<td>$622</td>
<td>$545</td>
<td>88%</td>
</tr>
<tr>
<td>Aggregated categories</td>
<td>$82</td>
<td>$21</td>
<td>26%</td>
</tr>
</tbody>
</table>


a. N=898

b. N=366

---

119. See generally Bovbjerg et al., supra note 94.
As Table 2 shows, there was a lot of missing data. The sample size for total verdicts was 898, but for non-economic damages it was only 366, an attrition rate of sixty-one percent. Two serious anomalies appear when we examine Table 2. First, for permanent serious injuries (Row 7 in Table 2) the median pain and suffering award was larger than the median total award, $1,642,000 versus $1,422,000. The pain and suffering component of the award cannot be larger than the total amount of the award, but for this level of severity the data show that it is 115%. Second, aggregating over all categories of severity, I calculate from Bovbjerg’s data that the median award for pain and suffering was twenty-six percent. Yet, in none of the disaggregated categories was the median as low as twenty-six percent; the figures ranged from twenty-nine to one hundred and fifteen percent. Neither of these findings makes sense. Perhaps an explanation for the oversight lies in the fact that the authors apparently aggregated over cases separately for total and non-economic awards and reported the results in two tables.120 The important point, however, is that if the missing data caused major anomalies in the median awards, we should view the other summary statistics—means, ranges, and standard deviations—with great skepticism as well.

2. The Dim Bright Line

Another problem is what I label the dim bright line. This oxymoronic label draws attention to the unwarranted assumption by researchers that a bright line separates economic from non-economic damages that allows them to reliably calculate the pain and suffering component of jury awards from verdict statistics. Economic damages may consist not only of past and present medical expenses and lost wages but also of things such as future medical expenses, future lost income, housekeeping expenses, modification of homes and automobiles, and costs for the education of dependent children. Non-economic costs are those that require a totally subjective judgment and for which there exists no metric for conversion to dollars, specifically, past and future pain and suffering, loss of life’s amenities, and loss of consortium. Often, all non-economic costs have been labeled simply “pain and suffering.” Should “disfigurement” be considered an economic or non-economic component, however, if it potentially affects employment or marriage potential? Is providing a surrogate for young children who have a brain-damaged parent an economic or non-economic injury? What are the economic implications of “loss of good

120. The proper calculation would be to find the difference between the total award and the estimated economic losses for each case and determine the medians of the remainders from these calculations.
offices and counsel?" It is not clear where one should draw the line between economic and non-economic components of damages.121

Most jurisdictions require only a general damage award.122 Neither verdict reporters nor closed-claim files contain the information that would allow disaggregation of these components. Ambiguity in classification could account for much variability in researchers' estimates of damage components from case to case. No bright line exists to separate the two components of damage awards, which adds substantial variability in the determination of the subtrahend, adding error variance in the remainder of the equation. If considerable unreliability exists in the criterion, it is not scientifically legitimate to ascribe the variability to jury behavior.

3. The Contentious Criterion Problem

Bovbjerg and his colleagues123 and Viscusi124 have drawn the conclusion that even though "pain and suffering" awards positively relate to seriousness of injury, great variability remains across cases within severity categories. From their data, Bovbjerg and his colleagues conclude: "[A]s one might expect of the least well-specified of all elements of damage, non-pecuniary awards show an even greater variability than do total valuations."125 Bovbjerg and his colleagues proceed to suggest remedies to assist the jury in its decisions.126

The difficulty with this interpretation lies partly in the missing data and dim bright line problems already identified. Even if there were no missing data and there existed a consensus on what constituted economic and non-economic damages, the problem of agreement on the "true" figures remains formidable, because the figures are still open to contention. Not only may the opposing parties disagree, but estimates for a single party may vary widely and there is no way to determine the "correct" figure.

Consider economic damages. Few observers of settlement negotiations or trials would agree with Peterson's assertions that economic damages can be "easily ascertained" and "usually . . . documented by routine business records."127 Bitter disputes between plaintiffs and defendants over past lost wages or other damages occur on a regular basis, and disagree-

121 See Viscusi, supra note 113, at 206 (making similar observation).
122 Some jurisdictions, e.g., New York, Florida, and Maryland, regularly use special verdicts in which juries specify the amount of the award for various components of the claim, including non-economic losses.
123 See generally Bovbjerg et al., supra note 94.
124 See Viscusi, supra note 30; Viscusi, supra note 113.
125 Bovbjerg et al., supra note 94, at 936.
126 Bovbjerg et al., supra note 94, at 936.
127 Peterson, supra note 111, at 9.
ments over future lost wages and medical costs may exceed disagreements about past economic loses by many orders of magnitude, particularly when attempts are made to reduce these costs to present value. One such case in the Duke Medical Malpractice Project involved a child with a severe birth injury; the parties disagreed about the costs of future medical and other health care costs for a child who could potentially live seventy more years as a blind, partially deaf, retarded, incontinent person susceptible to constant health problems. 128 Experts for the plaintiff produced a plausible case for an award exceeding $6 million. The defendant had obtained estimates from three separate experts that ranged from $2.1 million to $4.3 million. What was the "easily ascertained" correct figure? The same problem of valuation haunts cases of lesser magnitude. 129

When estimates of economic damages are obtained from insurers' closed-claim files or defense attorneys, the amount of estimated damages will be on the conservative side. 130 Similarly, when plaintiff lawyers provide the estimates, they will tend to be on the liberal side. The publishers of verdict reporters depend on the cooperation of attorneys to obtain their damage estimates. 131 Unlike social scientists who would be required to develop methodological standards for collection and coding of data, the publishers are under no professional strictures to use consistent methods in the gathering of statistics. If they obtain the damage estimate from plaintiffs in one case, defendants in another case, and they average estimates from both parties in another case, we would expect wide variability across cases due to the unreliability in the coding scheme. Similarly, we would expect wide variability if the awards are reported as actual values in some cases and "reduced to present value" in other cases, or if the awards are adjusted for comparative negligence in some cases but not in others. This variability poses a serious methodological problem for researchers, because the estimate of economic damages forms the subtrahend of the "pain and suffering" equation.

In short, the unreliability of the subtrahend figure may result from the wide variability in the estimates of jury "pain and suffering" awards, not from jury behavior. This alternative hypothesis is quite plausible given what we know about the unreliability of economic damage estimates and how researchers calculate pain and suffering damages. While it is most probably the case that there is unreliability in jury awards, 132 the percent-

128. See generally MEDICAL MALPRACTICE, supra note 1; Vidmar, supra note 50.
130. See generally DANZON, supra note 33; Viscusi, supra note 113.
131. See generally Daniels, supra note 1; Localio, supra note 6; Saks, supra note 1.
age of that unreliability cannot be calculated from verdict statistics absent some independent, reliable, and valid criterion of economic losses.

III. WHAT WE CAN LEARN FROM VERDICT STATISTICS

The discussion in Section II raises a number of serious methodological shortcomings of verdict statistics as data bearing on jury behavior. Verdict statistics can, however, provide important insights if used carefully, particularly if they are used in conjunction with other variables. Daniels and Martin's study of patterns of verdicts in medical malpractice cases in forty-two jurisdictions in eleven states indicated that jury trials did not uniformly yield high plaintiff-verdict rates or enormous damage awards.\textsuperscript{133} Similarly, I compiled data from twenty-one samples of malpractice cases collected by various researchers around the United States and found that plaintiffs prevailed in approximately thirty percent.\textsuperscript{134} The data in both of our studies are subject to the tip-of-the-iceberg problem, but the fact that plaintiff-verdict rates were consistently so low\textsuperscript{135} despite cross-jurisdiction differences appears to contradict the widespread claim that juries usually decide in favor of doctors.\textsuperscript{136} Daniels and Martin,\textsuperscript{137} Peterson and his colleagues,\textsuperscript{138} and Thomas Koenig and Michael Rustad\textsuperscript{139} have conducted research bearing on the frequency and magnitude of punitive damage awards. With respect to plaintiff-verdict rates, major questions of interest involved how often, in what types of cases, and in what amounts juries award punitive damages. The data provide important beginning answers about the scope of punitive damages, but I would caution that they do not tell us the "why" of damages, because they provide no information about the nature of trial evidence or juror reasoning. Danzon\textsuperscript{140} and Sloan and his colleagues\textsuperscript{141} compared statistics on jury verdicts with settlements. Such comparisons are subject to tip-of-the-iceberg and apples-and-oranges criticisms, but, if interpreted cautiously and with other data,

\textsuperscript{133} See Daniels & Martin, supra note 36, at 330, 340-41 (exhibiting plaintiff win rates and damage awards in table form).

\textsuperscript{134} See Data from Medical Malpractice, supra note 1, at 893 (stating plaintiffs' median 29.2\% of time).

\textsuperscript{135} See Data from Medical Malpractice, supra note 1, at 894 (setting forth chart of plaintiff win rates from numerous studies).

\textsuperscript{136} See Data from Medical Malpractice, supra note 1, at 893.

\textsuperscript{137} See generally Myth and Reality, supra note 44 (analyzing punitive damage awards overtime).

\textsuperscript{138} See generally M. PETERSON ET AL., PUNITIVE DAMAGES: EMPIRICAL FINDINGS (1987).


\textsuperscript{140} See generally DANZON, supra note 33.

\textsuperscript{141} SLOAN ET AL., supra note 98, at 153-77.
they provide useful insights and suggestive hypotheses for studies involving other types of research strategies. Theodore Eisenberg's research on plaintiff-verdict rates has led to insights bearing on George Priest and Benjamin Klein's hypothesis on the dynamics of how cases are selected for trial. Similarly, Samuel Gross and Kent Syverud cleverly used information about offers of settlement contained in verdict reports to test hypotheses about case selection and plaintiff-win rates.

Some studies have compared verdict data against other independent variables. Viscusi, Bovbjerg and his colleagues, Farber and White, and Taragin and his colleagues have compared jury awards with measures of the seriousness of the alleged injury. This is an acceptable use of verdict-statistics data because a number of sources have claimed that jury awards bear no relation to injury and that jury decisions on liability relate to seriousness of injury rather than negligence. Ironically, all of these studies show that jury damage awards positively relate to the seriousness of the injury at issue. Taragin and his colleagues' study further showed that verdicts on liability had no relationship to seriousness of injury. There has also been debate about whether jury verdicts on liability coincide with those potentially rendered by physicians. The studies by Taragin and his colleagues, Farber and White, and Sloan and his colleagues compared jury verdicts against assessments made by physicians acting in a neutral capacity. Sloan and his colleagues also compared damage awards against independent assessments of damages. The common factors in all of these studies are that the verdicts were compared to independent criteria on a case-by-case basis, and that they tested hypotheses that are independent of how cases are selected for trial. Consequently, the hypotheses can be tested relatively free of confounding factors.

---


143. See generally Gross & Syverud, supra note 37.

144. See generally Bovbjerg et al., supra note 94; H. Farber & M. White, Medical Malpractice: An Empirical Examination of the Litigation Process, 22 RAND J. OF ECON. 199 (1991); M. Taragin et al., The Influence of Standard of Care and Severity of Injury on the Resolution of Medical Malpractice Claims, 117 ANNALS OF INTERNAL MED. 780 (1992); Viscusi, supra note 113.

145. See generally AM. MEDICAL ASS'N, supra note 46; MANHATTAN INST. FOR POLICY RESEARCH, supra note 19.

146. See generally AM. MEDICAL ASS'N, supra note 46; PHYSICIAN PAYMENT REVIEW COMM'N, supra note 55; U.S. DEP'T OF JUSTICE, REPORT OF THE TORT POLICY WORKING GROUP ON THE CAUSES, EXTENT, AND POLICY IMPLICATIONS OF THE CURRENT CRISIS IN INSURANCE AVAILABILITY AND AFFORDABILITY 63 (1986); Johnson et al., supra note 56; L. Tancredi, Compensating for Medical Injuries, 1986 N.Y. ST. J. OF MED. 370.

147. See generally SLOAN ET AL., supra note 98; Farber & White, supra note 144; Taragin et al., supra note 144.

148. See generally SLOAN ET AL., supra note 98; Farber & White, supra note 144; Taragin et al., supra note 144.

149. Yet, even here I cannot resist offering another methodological admonition. The reliability of
The above list of studies is not exclusive and is intended only to illustrate that jury verdict data can provide highly useful information about jury behavior. The studies, however, must accompany a careful interpretation and recognition of the methodological problems identified in this article.

IV. SUMMARY AND CONCLUSIONS

The jury plays a key and highly visible role in the American tort system. Those persons and groups dissatisfied with the tort system have long blamed its perceived ills on juries. In the past, they based their critiques upon anecdotes and appeals to intuition that a body of laypersons cannot decide cases as competently as trained judges. The anecdotes always raise the problem of how representative the cases are, but, in addition, the anecdotes often turn out to be misleading or, in some instances, fabrications.150

While appeals to intuition and anecdotes remain part of the arsenal of jury critics, these critics in recent years have also drawn upon studies of verdict statistics conducted by social scientists. These studies are beguiling because they involve data on hundreds or even thousands of actual jury outcomes. Moreover, the results of many of the studies superficially appear to support what the critics claim to have known all along: that juries favor plaintiffs in "deep pockets" cases; awards are spiraling; verdicts are capricious or biased; and pain and suffering constitutes the bulk of awards. I have shown in this article that many of the conclusions drawn from these studies have no scientific validity. The methodological limitations of the databases and errors in conceptualizing the issues do not allow such conclusions. There are plausible alternative explanations of the relationships that have been found. Social scientists who have conducted the research have overlooked many of these methodological problems and legal scholars, judges, and other commentators have accepted the data without criticism because the conclusions correspond with their pre-existing biases.

It is essential to point out that my analysis of these verdict-statistics studies does not prove the opposite conclusion, namely that juries perform

150. See generally supra notes 15-61 and accompanying text. See also Marc Galanter, Pick a Number, Any Number, AM. LAWYER, Apr. 1992, at 82.
their tasks competently. It only says that by themselves the statistics cannot tell us one way or the other. I have offered the view that verdict statistics can shed light on certain issues in the debate about what juries do, particularly when they are analyzed in conjunction with other variables. Even these studies, however, give no insights into why juries decide as they do, because verdict statistics tell us nothing about the minds of jurors or about the quality and quantum of evidence they hear at trial.

While my goal in this article has been confined to the methodological critique of verdict-statistics studies, I believe it is important to mention that recent research using other methodologies and databases in addition to verdict statistics contradicts the claims that juries are incompetent and biased. For example, I conducted research on medical malpractice juries using multiple methods and data sources. In a series of articles and a book, I have shown that, on the whole, juries do not favor claimants over doctors, do not make negligence judgments based on the depth of defendants' pockets or the severity of patients' injuries, and that their verdicts on liability are actually consistent with doctors' judgments of negligence. On the issue of damages, the data have shown that awards generally relate to the seriousness of injury, that variability in awards can probably be ascribed to the jury responding to the trial evidence, that pain and suffering does not constitute the vast proportion of awards, that juries can render pain and suffering damages more reliably than judges, and that punitive damages are rare in malpractice cases. Studies by researchers dealing with civil juries in other types of cases have tended to support the jury system as well. These various findings appear to exonerate the tort jury of some of the most extreme charges against it, and, in most respects, give it a favorable grade for performance of its assigned tasks.

Empirical issues are at the heart of the debate about the role and performance of juries in the tort system. Empirical data are needed, but jury verdict statistics can shed light on only limited aspects of the debate. Whether it is verdict statistics or other forms of data, however, we must separate pap from circumstance. In this article, I have provided some conceptual tools that can aid in this task.

151. See generally Empirical Evidence, supra note 1; Medical Malpractice, supra note 1; Vidmar, supra note 50, Vidmar & Rice, supra note 132.
152. See generally Medical Malpractice, supra note 1.