

Underclaiming and Overclaiming

Sachin S. Pandya and Peter Siegelman

Arguments that we have too much litigation (overclaiming) or too little (underclaiming) cannot be valid without estimating how many of the undecided claims that are brought (actual claims) or not brought (potential claims) have or lack legal merit. We identify the basic conceptual structure of such underclaiming and overclaiming arguments, which entails inferences about the distribution of actual or potential claims by their probability of success on the merits within a claims-processing institution. We then survey the available methods for estimating claim merit.

I. INTRODUCTION

In 1995, four state attorneys general sent a letter to the *New York Times* to champion legislation making it harder for prisoners to sue (Vacco et al. 1995). They argued that prisoners, with “plenty of time and plenty of grudges,” often filed meritless lawsuits. As proof, they cited three prisoner suits as “typical”: a suit alleging “cruel and unusual punishment” for receiving a jar of creamy peanut butter, not chunky as requested from the prison canteen; a suit complaining of “no salad bars or brunches on weekends and holidays”; and a suit filed because the inmate wanted beige prison towels rather than white ones. The attorneys general also pointed to a study of (non-habeas) inmate litigation, which found that 94 percent of prisoners who file suit “win absolutely nothing.”

Aside from misrepresenting the allegations in the lawsuits they cited (Newman 1996, 520–22), the argument of the attorneys general had an important flaw. It did not address the merits of legal claims that prisoners could have brought but did not bring (their potential claims). For example, while the attorneys general conceded that cases based on “sexual assault by prison guards . . . should not be dismissed or disregarded as nonmeritorious,” they did not address whether prisoners could and did actually bring claims based on such assaults, or whether the prison litigation reform they backed would make it harder to do so. (A 2008–2009 survey of inmates in prisons and jails estimated that over 88,000 prison and jail inmates nationwide had suffered some form of sexual victimization in the past twelve months [Beck and Harrison 2010, 6].) Even if many actual prisoner lawsuits lack merit, that would not necessarily justify making it more difficult for prisoners to sue if such reforms would also discourage meritorious prisoner lawsuits.

In 2008, two law professors made a related mistake. Brake and Grossman (2008, 861) declared Title VII, the leading federal employment discrimination statute, a

Sachin S. Pandya is an Associate Professor of Law at the University of Connecticut School of Law, Hartford, Connecticut.

Peter Siegelman is the Roger Sherman Professor of Law at the University of Connecticut School of Law, Hartford, Connecticut.

“failure as a rights claiming system,” largely because Title VII’s time-to-file rules and rules against employer retaliation are too stringent, given evidence that most people have a hard time perceiving or reporting themselves as having suffered discrimination. Brake and Grossman’s mistake was to assume, without proof, that such rules primarily discourage or preclude *otherwise meritorious* Title VII claims. People often hold erroneous beliefs about the relevant law or the facts, making it unreliable to infer a potential claim’s merits from the potential claimant’s *perceptions*. For example, many workers believe that the law allows the firing of an employee only if the employer has a good reason (“good cause”)—even in jurisdictions where that belief is false. (For a review of this literature, see Roehling [2002].) Similarly, even if people perceived themselves as experiencing employment discrimination, those perceptions will not always match what a court would decide, had those people brought and litigated their claims.

These examples concern errors that sometimes appear in a class of arguments about how well courts, administrative agencies, arbitral forums, and other claims-processing institutions (including informal grievance resolution inside firms) encourage claims with merit and discourage those without it. We call these *underclaiming* and *overclaiming* arguments. These arguments appeal to a particular view of how to judge a claims-processing institution. We should not care about the raw number of actual or potential claims that a claims-processing institution did or could have decided. We should care only (or, at least, mostly) about those claims that, if brought and decided, would have prevailed on their merits. If there are many perceived injuries, and thus potential claims, but most would fail on the merits if brought, we should be glad that those potential claims were never brought. Conversely, we should worry if many potential claims with merit were not brought and decided, even if many meritless claims were. The problem is that many underclaiming and overclaiming arguments assume, assert, or imply, rather than prove, the merits of (actual or potential) undecided claims—that is, how these claims would have fared had they been brought and decided.

This article identifies the conceptual structure of valid underclaiming and overclaiming arguments. Such arguments entail inferences about the distribution of actual or potential claims by their probability of satisfying the rules for deciding claims within a particular claims-processing institution. The article then canvasses five research strategies for estimating such probabilities, as illustrated with examples from research on several types of claims, including employment discrimination, medical malpractice, benefit applications, and insurance fraud. In so doing, the article contributes to the subset of sociolegal research that is motivated by underclaiming or overclaiming concerns, such as research into unmet “legal needs” (A. Currie 2006; Legal Services Corporation 2009) and, as the above examples suggest, much of modern policy discourse about civil litigation.

As recent reviews of the sociolegal literature (Sandefur 2008; Kritzer 2011) indicate by omission, this literature has not fully identified the unique analytic structure of underclaiming and overclaiming arguments or the set of empirical methods appropriate for testing them. Some prior work has recognized the need to consider underclaims and overclaims simultaneously. For example, Saks (1992, 1169) motivated his review of data on the U.S. tort system by asking how well that system compensated plaintiffs with “actionable” injuries, including how much it failed to pay plaintiffs for “actionable” injuries (underclaims) and how much it paid plaintiffs without “actionable” injuries (overclaims).

Similarly, when some (e.g., Priest 1987) blamed excessive tort litigation for sharp spikes in liability insurance premiums, Abel (1987, 447) responded by declaring that the real tort “crisis” was a “crisis of underclaiming rather than overclaiming,” in part because many “tort victims” fail to claim. Proving this, he acknowledged, required “establish[ing] the population of injury victims who *could* recover tort damages before examining how many of them, and which ones, actually do so” (448). In contrast, Brickman (2011, 120) concluded that “the concept of the meritorious claim is dubious at best,” either because it requires estimates of a tort claim’s “probability of success based on previous jury verdicts,” which are “highly variable” and sometimes “defy rationality,” or because judging a particular claim’s merit is an irreducibly normative issue, requiring “widespread agreement” on a set of policy goals and how well any decision on that particular claim advances those goals.

Our chief contribution is to show precisely how underclaiming and overclaiming *can* be objects of sociolegal study. This kind of work lies at the border between sociological research on disputing (Felstiner, Abel, and Sarat 1980) and traditional legal scholarship that aims to clarify the content of legal rules on the books. To assess the merits of undecided claims, actual or potential, one must predict how a claims-processing institution would decide them, given its decision rules. For legal institutions, this sort of analysis sounds a lot like what lawyers are trained to do. In fact, however, sociolegal research can contribute much to such analysis. Only sociolegal research can help answer questions such as “Why do potential plaintiffs with meritorious claims fail to bring them?” and “Under what circumstances do plaintiffs bring claims that lack merit?” Ultimately, we hope to blur the boundaries between sociolegal research and traditional legal analysis because both are necessary to understand how well the law works in practice.

II. UNDERCLAIMING AND OVERCLAIMING ARGUMENTS

Underclaiming and overclaiming arguments entail inferences about the distribution of actual or potential claims by their probability of success on the merits within a claims-processing institution. By “on the merits,” we mean the probability that a claim, if brought, would satisfy a particular claims-processing institution’s own rules for deciding claims.

Figure 1 illustrates the basic intuition. Let I = the set of perceived injuries, C = the set of claims actually brought, and M = the set of claims (actual or potential) with merit. Underclaims are claims that, if they had been brought, would likely have succeeded on the merits ($= (I \cap M) - C$). Overclaims are actual claims that lack merit. As Figure 1 depicts, underclaims and overclaims can exist simultaneously. This means that every argument about the extent of underclaiming is incomplete unless it also estimates the number of overclaims, and vice-versa.

We can further identify two types of overclaims. Overclaims₁ are actual meritless claims where the claimant perceived injury ($= I \cap \sim M \cap C$). Overclaims₂ are actual meritless claims, those for which the claimant perceived no injury ($= \sim I \cap C$) but nevertheless made a claim. By “injury,” we mean simply an event that, under the decision rules of the claims-processing institution, must have occurred for the claimant

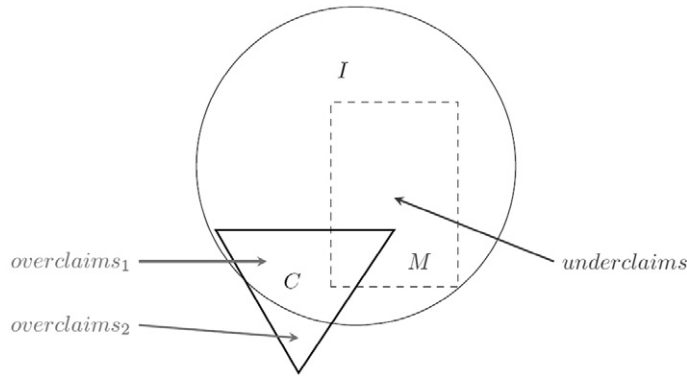


FIGURE 1.
Venn Diagram of Overclaiming and Underclaiming Arguments

to be eligible for a favorable decision. Every claimant asserts “injury” in this sense, and the claims-processing institution must decide whether, under its decision rules, an “injury” in fact occurred. For this reason, Figure 1 reflects a simplifying assumption: any person who does not perceive an injury but claims that he or she has suffered an injury has brought a claim that is unlikely to succeed on the merits.

This conception of success on the merits has two major implications. First, a claim (actual or potential) has merit only with respect to a particular institution’s rules for deciding claims. Because those rules vary by institution and time, a claim’s merit necessarily varies by institution and time as well. Second, evaluating claims-processing institutions requires assigning social values to each underclaim and overclaim. We consider each implication in turn. We then contrast our approach with the dispute-pyramid model. Despite that model’s influence on sociolegal research, it does not account for claim merit, and therefore cannot properly be used to test underclaiming or overclaiming arguments.

A. Institutions

A claim (actual or potential) has merit only with respect to a particular claims-processing institution’s own rules for deciding claims. Claims-processing institutions include not only courts, but also administrative agencies, commercial arbitration, workplace grievance systems, and insurance claims departments, among others. For every actual or potential claim, whether that claim has merit or not is a function of a set of decision rules, which vary by institution and time.

For every claim that a particular claims-processing institution actually decides, we can observe the decision, and thus whether the claim has or lacks merit. For every potential claim that this institution did not decide, however, we have to estimate that claim’s probability of success *as if* that institution *had* decided it. For ease of exposition, we call every undecided claim *c* (actual or potential) an overclaim if that claim was actually brought and its probability of success is less than or equal to 50 percent

($p_c \leq 0.50$). We call that claim an underclaim, whether actually brought or not, if its probability of success is more than 50 percent ($p_c > 0.50$). (We might easily have adopted different thresholds. We could have said, for example, that claim c is an underclaim if $p_c > 0.75$.)

This definition indexes claim merit to an institution's *own* decision rules,¹ because underclaiming or overclaiming arguments usually attempt to sidestep disagreements about the normative premises or implications of an institution's decision rules, instead taking those rules as objective givens. In this respect, our definition differs from those that treat claim merit as a function of researcher-created (external) decision rules that are institution independent. For example, in some studies of closed medical malpractice claims filed with liability insurers, the researchers (Gandhi et al. 2006, 490; Studdert et al. 2006, 2026; Studdert and Mello 2007, S56) assessed claim merit not by reference to a state's tort law governing negligence, but by using the Institute of Medicine's definition of "error": whether the adverse event suffered by the claimant was caused by "the failure of a planned action to be completed as intended (i.e. error of execution) or the use of a wrong plan to achieve an aim (i.e. error of planning)." The researchers stressed to their claims-file reviewers that this definition of error is "not synonymous with the legal definition of negligence and that a mix of factors extrinsic to merit influence whether claims are paid during litigation."

This general approach—assessing claim merit using external criteria—has several advantages. First, it allows researchers to compare actual or potential claim merit across institutions with different decision rules. Second, for claims actually brought and decided by an institution, the researcher can judge a claim's outcome based on how well it accords with the researcher's external criteria. Thus, for example, one can treat actual court decisions on medical malpractice claims as "wrong" if "wrong" is defined as the failure of those decisions to match independent physician assessments of whether those claims satisfy the Institute of Medicine's definition of medical error.

In contrast, by indexing claim merit exclusively to a claims-processing institution's application of its own decision rules, the researcher is largely precluded from treating that institution's actual decision on a claim's merits as "wrong." Moreover, where a particular undecided claim has a nonzero probability of success on the merits in multiple claims-processing institutions, there is no *single* claim-specific probability of success on the merits; rather, a claim could be an underclaim with respect to one institution (for which $p_c > 0.50$), but not with respect to another institution (for which $p_c \leq 0.50$). To be sure, once the researcher elects to study a particular institution, he or she can simply ignore what that claim's probability of success is in other institutions.

In practice, however, many underclaiming and overclaiming arguments adopt or imply a conception of merit indexed to a particular institution's decision rules *and* some external criteria. For example, for every actual or potential claim c , the researcher, in estimating the probability of the claim's success on the merits in a particular institution, may choose to *ignore* the subset of that institution's decision rules that count as "procedural." (In civil courts, such rules include statutes of limitation, filing deadlines, administrative exhaustion requirements, and rules governing finality of judgments.) To

1. These decision rules include those for determining whether the claimant has suffered an "injury" at all.

identify that subset, a researcher can follow the institution's own rules for identifying the subset of "procedural" rules,² or rely on some *external* criteria for what counts as a "procedural" rule.

To illustrate, suppose a potential claim counts as an underclaim ($p_c > 0.50$) only if the researcher chooses to ignore the claims-processing institution's rule under which the time to file that claim has already expired. If the researcher does not ignore this rule, the potential claim's probability of success drops to zero ($p_c = 0$), and thus would not be counted as an underclaim. To decide which probability corresponds to that claim, we must decide whether or how to identify the institution's "procedural" rules. That in turn depends on the goals of the research. For example, it might be appropriate to ignore that time-to-file "procedural" rule, and treat the claim as an underclaim, if the research aims to find out how shorter or longer statutes of limitation affect claims-processing outcomes.

Similarly, many underclaiming and overclaiming arguments also implicitly adopt external criteria for estimating claim merit by assuming that the claims-processing institution under study has certain features that in fact it may not have, such as an equal opportunity to present evidence, decisionmaker impartiality, or a stable norm of treating like claims alike. Underclaiming or overclaiming estimates will be contested based on how much disagreement there is over whether such features are in fact present and, if not, whether they *should be*, that is, whether their presence is necessary for accurate estimation of claim merit.

As an example, consider the operation of mandatory arbitration in employment disputes. Critics have complained that arbitrators who are involved in more than one case with a given employer are biased in favor of that employer. For evidence that the critics may be right, see Colvin (2011, 15, 18). If repeat-play arbitrators are in fact biased in this fashion, and if everyone agrees that, absent decisionmaker impartiality, past arbitrator decisions should not count as decisions on the merits, then estimates of claim merit based on those past arbitrator decisions will be contested as systematically biased downward. This implies that in order to properly estimate claim merit for employment arbitration, the researcher must estimate the probability of success on the merits in a way that not only accounts for the decision rules in employment arbitration, but also corrects for arbitrator bias against workers.

B. Claim Values

Underclaims and overclaims may vary in the value that society attaches to each. The costs of an underclaim may include the foregone relief to people eligible for it, or the reduction in future harmful behavior (deterrence), if any, that would have occurred had the claim been made. Each overclaim may impose a social cost due to wasted adjudicative resources, undeserved settlements, or, in some cases, taking away funds

2. For example, since Congress has prohibited the U.S. Supreme Court from prescribing rules of procedure that "abridge, enlarge or modify any substantive right," 28 U.S.C. § 2072(b), the Court has had to decide in many cases whether a prescribed rule is procedural or substantive, see *Shady Grove Orthopedic Associates, P.A. v. Allstate Insurance Co.*, 130 S. Ct. 1431, 1442–43 (2010) (collecting cases).

TABLE 1.
Hypothetical Claiming Example

	Merit ($p_c > 0.50$)	Meritless ($p_c \leq 0.50$)	Total
Perceived injuries	200	490	690
Claims	30	20	50
Claiming rate	15%	4.1%	7.2%

available to pay those who *do* have claims with merit.³ Indeed, much as some have expressed tolerances for criminal justice system outcomes in terms of the ratio of innocent persons jailed to guilty persons freed, we can judge claims-processing institutions in terms of tolerable shares of underclaims and overclaims.

Yet, absent agreement on how to make the costs of underclaims or overclaims commensurable, there is no justification for subtracting overclaims from underclaims (or vice versa) in order to measure whether a claims-processing institution has too much or not enough claiming overall.

To illustrate, consider the stylized example set out in Table 1. Suppose that, in a sample of 10,000 people, 200 of them have a potential legal claim with merit ($p_c > 0.50$). Suppose further that while all 200 of them perceive the injury associated with their (potential) claim, only thirty (15 percent) actually make the claim, and go on to prevail on the merits. Nobody else (the remaining 9,800 people) has a potential legal claim with merit (so for all c in this remaining group, $p_c \leq 0.50$). However, 5 percent of this group (490) misperceive that they have a legally cognizable injury, and twenty of them go on to make a (meritless) claim. None of their actual claims succeeds on the merits. The end result is that there are fifty total claims and 690 perceived injuries, for a raw claiming rate of 7.2 percent. Conditional on a perceived injury, the claiming rate for persons with a potential claim with merit is 15 percent ($30/200$), while the claiming rate for those whose (potential) claims lack merit is 4.1 percent ($20/490$), less than one-third as large. This system has twenty overclaims (all overclaims₁) and 170 underclaims.

Although the share of meritless claims among all claims is $20/50 = 40$ percent (which, taken alone, suggests substantial overclaiming), the proportion of meritless potential claims is only $490/10,000 = 4.9$ percent, which suggests negligible overclaiming. Yet without prior agreement as to the social costs of underclaims and overclaims, we cannot know whether we have too much or too little claiming overall. In the example, 40 percent of actual claims lacked merit ($20/50$), but 85 percent of potential claims with merit ($170/200$) were not brought. (Unfortunately, this problem exists even if we

3. Policymakers might also infer from a large number of overclaims₁ that the applicable decision rules do not match popular views of what those rules should be, which in turn may reduce the legitimacy of institution outcomes. For example, the fact that many workers believe they have experienced discrimination could be a problem, even if most of the discrimination claims actually brought by those workers would likely fail on the merits. Such a mismatch, however, is not a social cost of overclaiming itself.

measure claim merit continuously by letting the value of each claim vary with its probability of success. We discuss this issue in more detail in the Appendix.)

Finally, attention to claim valuation resolves any complication that arises if a claim is not decided by the claims-processing institution under study, but *has* been decided on the merits by *another* claims-processing institution. To illustrate, suppose that a worker's claim of race discrimination in hiring has a nonzero probability of success in multiple institutions (e.g., state court, federal court, private arbitration), each with its own (sometimes partially shared) rules of decision. The researcher estimates that, in federal court (the claims-processing institution of interest), the claim's probability of success on the merits exceeds 50 percent, and thus, for that institution, that claim counts as an underclaim ($p_c > 0.50$). The researcher then learns that, in fact, the potential claim in question had actually been brought *and decided* in state court in the worker's favor (as, say, a violation of a state statute). This claim is by definition not an underclaim with respect to a state court because it *was* brought and decided in that state court. The question is this: Given this new information, should the researcher still count that same potential claim (i.e., race discrimination in hiring) as an underclaim with respect to federal court?

The answer is that the researcher should continue to count that potential claim as an underclaim with respect to federal court, but adjust the *value* of that underclaim to reflect the degree to which society would have been better off had that potential claim been decided in the institution under study (here, federal court). If society should be indifferent as to whether the claim was decided on the merits in federal court or state court, *and* if both outcomes would have been the same in any event, then we should treat the underclaim as reflecting *zero* social loss. In contrast, by not counting it as an underclaim at all with respect to federal court, the researcher makes a claim value judgment implicitly that should be made explicitly.

C. The Dispute-Pyramid Model

Underclaiming and overclaiming arguments are distinct from the influential dispute-pyramid model because that model does not purport to estimate the probability of success of an actual or potential claim on the merits.

Developed in the early 1980s in the United States as part of the Civil Litigation Research Project (CLRP) (Felstiner, Abel, and Sarat 1980; Miller and Sarat 1980), and used in other countries (see Harris et al. 1984; Genn 1999; Murayama 2007), the model represents the disputing process in terms of a population pyramid. While every injurious experience (as defined by the researcher) could potentially lead to a lawsuit, only some of those who are injured perceive that injury; only a subset of this group blame someone for their condition; only a subset of those who blame someone for their condition actively seek redress; only a subset of those who actively seek redress seek out a lawyer; and only a subset of those who seek out a lawyer ever file a lawsuit or make a formal claim of any kind.

The key fact about the dispute pyramid is that it narrows from a broad base of potential claims to a much smaller peak of filed cases. For example, consider the set of people deliberately injured by their spouses. If most of them believe that they, not their

spouses, are to blame for their injuries, any private right of action for domestic violence victims might only negligibly affect abuse rates because the plaintiffs needed to enforce this protection never materialize. Since they blame themselves, they do not bring claims against their spouses, and they fail to move up the dispute pyramid from some early stage before formal litigation would have occurred.

With this claiming model, CLRP took the “dispute” or “case” as its unit of analysis and compared disputes across several dimensions, including type of dispute, type of disputant (individuals, organizations), and type of dispute-resolution institution (courts, agencies, private arbitration, resolution without resort to a third party). In the CLRP household survey, for example, respondents were asked if they had recently experienced one of several predefined kinds of events. Then, in follow-up interviews, those respondents were asked how they behaved, including, but not limited to, doing nothing at all, talking to a lawyer, or filing a lawsuit.

By asking about claiming behavior conditional on a self-reported perceived event, this approach captures potential disputes that would be missed by looking only at court opinions or dockets. On the other hand, this approach omits overclaims₂, which by definition occur absent any perceived injurious event. This approach also cannot capture unperceived events that, because they are unperceived, cannot be self-reported in surveys. For example, job applicants rarely know that they have not been hired because of their race or sex. Yet, though legally prohibited, employer surveys and audit studies suggest that employers commonly do hire on the basis of race and sex (Bertrand and Mullainathan 2004; Moss and Tilly 2006). Unperceived events may also include bodily harms that are not immediately apparent. If years pass between the time a company dumps toxic chemicals and the time a person who lives near the dump site can be accurately diagnosed with cancer, that person will be less likely to attribute his or her cancer to the dumped chemicals, even if the dumping caused his or her cancer.

Most importantly, the dispute-pyramid model does not purport to help estimate the probability of success on the merits of a potential claim. Nonetheless, some have based overclaiming or underclaiming arguments in part on this kind of research. For example, in arguing that many “tort victims” fail to claim, Abel (1987, 449–50) relied on, among other studies, Harris et al. (1984), even though that study, following a variant of the CLRP research design, did not in fact try to assess the legal merits of potential claims arising from its respondents’ self-reported injuries. While the dispute-pyramid model is a valuable tool for understanding the way institutions process disputes, since that model does not estimate claim merits, it is inappropriate for testing any underclaiming and overclaiming arguments that are central to much policy analysis.

III. EMPIRICAL STRATEGIES

The rest of this article sketches the range of available empirical strategies for estimating a potential claim’s probability of success on the merits. With an unlimited research budget, we could pay teams of private investigators to fan out and build, for every self-reported injury, a dossier that provides perfect information about the actual circumstances of the self-reported injury. In turn, that dossier could be handed over to

panels of independent legal experts to score the claim's odds of success in court if it were brought as a legal claim under every available legal theory. (Even this research design would not capture unperceived injuries, or claims made absent perceived injuries, as, for example, fraudulent claims.)

In the real world of limited budgets and human subject research protocols, there are five basic strategies for estimating the degree of under- or overclaiming in a particular claims-processing institution: (1) retrospective estimates by experts; (2) contemporaneous estimates by situated actors; (3) inference from observable claim characteristics; (4) field experiments or audit studies; and (5) inference from the characteristics of filed claims. Each strategy starts with a conception of merit indexed to a particular institution's decision rules, and further assumes a set of institution- and time-invariant background conditions.

A. Retrospective Estimates by Expert Raters

In this strategy, researchers obtain access to routinely collected documents relevant to a potential or actual claim, use independent experts to review those documents to estimate claim merit, and score the expert's confidence in that estimate.

For an example of this approach, consider the Harvard Medical Practice Study (HMPS) (Harvard Medical Practice Study 1990; Brennan et al. 1991; Localio et al. 1991). The HMPS researchers randomly sampled the medical records of patients discharged in 1984 from hospitals in New York State. Medical record administrators initially screened the sampled patient records. For each patient record that met one or more screening criteria, two physicians independently reviewed that patient's record and scored it on a six-point scale based on the reviewer's confidence that (1) the patient had suffered an adverse event caused by the medical care provided, and (2) that such care amounted to negligence.

The HMPS researchers operationalized negligence by asking their physician-reviewers, for cases with adverse events caused by the medical care provided, whether that event "was possibly due to a reasonably avoidable error, or carelessness by either an individual or medical care system, or both." Reviewers were also asked to consider seven mitigating or aggravating circumstances, including the "degree of deviation of treatment from accepted norms," case complexity, and degree of "emergency in management of case prior to occurrence of negligence." They were then asked again whether there was negligence. If they thought that there was, they were asked to rate its severity (slight, moderate, or grave) and estimate their level of "confidence in the evidence of negligence" (Harvard Medical Practice Study, Technical App. 5.IV.1, Adverse Event Analysis Form 1990, 10, 13–14).

By this method, the HMPS researchers identified 1,133 adverse events that occurred during patient hospitalization (or during an earlier hospitalization or outpatient care and discovered during that hospitalization). They judged 280 of these to have resulted from negligence. After adding sample weights, the HMPS researchers estimated that 98,609 (3.7 percent) of the patients discharged from hospitals in New York in 1984 suffered adverse events, and 27,179 of them (27.6 percent) were due to negligence (Brennan et al. 1991, 371).

The HMPS researchers then matched patients in the sample with medical malpractice claims filed with liability insurers and self-insured hospitals. They found only forty-seven medical malpractice claims, from which they estimated that only 415 of the adverse events caused by negligence (1.53 percent) led to a claim filed (Localio et al. 1991, 248). A follow-up record review found four additional claims, for a total of fifty-one medical malpractice claims (Brennan, Sox, and Burstin 1996). This was strong evidence of medical malpractice underclaiming.

Moreover, thirty-nine of the forty-seven claims originally found were ones for which the HMPS physician-reviewers found no evidence of an adverse event caused by negligence. Although this was a basis for inferring medical malpractice overclaiming, it was only a weak one, in part because of the small sample of actual claims and because HMPS-reviewer protocols for finding negligence did not precisely match the rules governing how negligence and causation would be found in civil litigation (Localio et al. 1991, 249; Baker 2005, 402–506). Although other studies of medical errors and claiming have varied in their findings, what matters here is that the Harvard Medical Practice Study illustrates the approach of using expert raters to estimate actual and potential claim merit.

For this approach to be useful, two important conditions must be satisfied. First, there must be a close match between the relevant rules of decision of the claims-processing institution (e.g., the legal rules that a court would apply) and the criteria with which expert raters judge claim merit. The HMPS researchers attempted to create such a match, as did the Utah and Colorado Medical Practice Study, a similarly designed study of hospital records in Utah and Colorado (see Studdert et al. 2000; Thomas et al. 2000). That study defined “negligent adverse events . . . , in accordance with standard tort criteria, as actual injuries proximately resulting from a treating physician’s failure to meet the standard expected in his practice community” (Studdert et al. 2000, 252). Similarly, Cranberg, Glick, and Sato (2007, 156) judged closed medical malpractice claims files concerning neurologists in the light of “the standard of tort law,” that is, whether significant patient harm was caused by neurologic care “below the standard that would be expected of any neurologist under similar circumstances.” In their study of closed medical malpractice claim files for anesthesia-related patient injuries, Cheney et al. (1989, 1599) judged the “overall appropriateness of anesthetic care . . . as ‘that which met the standard of care for a prudent anesthesiologist practicing anywhere in the United States at the time of the event’” and judged inappropriate care “as ‘that below the standard of practice (i.e., negligence)’.”

The validity of using expert raters depends in part on how closely such raters agree with one another. To test the reliability of physician-rater judgments in the Utah and Colorado Medical Practice Study, Thomas et al. took separate random samples of the hospital records that, in that study, physician-raters had found to indicate “no adverse event,” “adverse event,” or “negligent adverse event.” They then subjected the sampled records to a second and third physician-rater review. They found “moderate to poor inter-rater reliability” (2002, 814). Adverse event rates were “markedly different” depending on the minimum level of rater confidence necessary to indicate the presence of an adverse event, or depending on the required minimum number of raters who detected that event (814). For example, of the 161 files originally found to indicate a “negligent adverse event,” a random sample of fifty was subjected to a second and third

review (815). When the researchers set the minimum rater confidence score at 2 (out of 6), the negligent adverse event rate was 32.21 percent. That rate dropped to 15.13 percent when the minimum rater confidence score rose to 4. Requiring two of the three raters to agree on a rating further pushed down the adverse event rate to only 4 percent (814, Table 2).

In addition to the problem of interrater reliability, the expert-rater approach requires a stable infrastructure of reliable record keeping in which data are routinely kept for ordinary behavior, regardless of whether such behavior leads to any actual claiming. Hospitals, for example, keep voluminous records of patient treatment and health for *all* patients (not just those whose treatments they mishandle in some way). Studies such as the HMPS thus depend on modern routines of making and preserving medical records, the result of a shift to and standardization of the hospital patient case file in the early twentieth century (Craig 1989–1990, 1990; Berg and Haterink 2004). Few settings other than hospitals have a comparable infrastructure for producing reliable documentary evidence of both perceived and unperceived injuries. For example, while any accidental injury motivating a hospital visit will generate records of both injury and treatment decisions, it will not necessarily produce a police report. And even if it does, police report classification of, for example, injury severity does not perfectly match injury severity measures based on hospital records (Sciortino et al. 2005; Amoros et al. 2007; Tsui et al. 2009, 86), suggesting that police reports are less reliable than hospital records for studying how injury severity affects underclaiming of accidental injuries.

Similarly, personnel files do not contain a complete record for every legally relevant workplace event (e.g., a firing, promotion denial, or refusal to hire); neither do they document what motivated the manager or co-worker responsible for the event. Thus, where the issue is whether a person was fired not because of poor job performance, but because of discrimination or other impermissible motives, personnel files will rarely provide direct evidence of manager motivations or states of mind. Accordingly, since the manager's motives are crucial to constructing a legal case for discrimination, personnel records alone are usually inadequate for assessing the merits of a worker's discrimination claim. By contrast, the evidence required to establish medical negligence *does* routinely appear in medical records.

The absence of relevant evidence makes causal inference (and thus claim merit estimation) more difficult. For example, Zwerling and Silver used an employer's complete personnel records to estimate a model of the factors leading to employee firing. Even after controlling for all variables in the employer's files, they concluded that there was a racial disparity in firing rates. However, when they asked an independent human resources manager to review the files, he found no "inappropriate" firings in the records he reviewed, though eleven firings presented "unclear documentation for the precise reason for the discharge" (1992, 658). Zwerling and Silver suggested that a kind of favoritism might partly explain their results: all firings were legitimate, regardless of race, but some whites were *not* fired for offenses that would have led to their termination had they been black. While this inference is plausible, the problem is that the files on which the authors relied did not contain evidence of *all* workers' behavior, so it is impossible to know whether there were in fact white workers who misbehaved but whose offenses were never recorded in their files.

B. Contemporaneous Estimates by Situated Actors

Another strategy for estimating claim merits relies on documents containing the contemporaneous estimates of claim merit by actors who routinely make such judgments and, in theory, have strong incentives to make accurate estimates. Accordingly, liability insurance claims files should contain claims adjuster estimates of claim settlement costs. The files of a plaintiff's lawyer may contain an estimate of the odds of the potential lawsuit's success. Administrative agency staff files of citizen complaints may contain staff estimates of merits for purposes of deciding which complaints to investigate or litigate.

Using this strategy to evaluate claim merits has several drawbacks. First, it will underestimate underclaims. A person needs to have already complained to attract the attention of the merit evaluators. Second, in predicting claim merits, these actors may consider other factors that the researcher may want to treat as irrelevant to claim merits (e.g., the effect of "procedural" rules).

Third, some institutional actors estimate merit based on a claim's expected monetary value, without expressly disaggregating odds of liability from expected damages. So, for example, a claims adjuster may indicate in a claims file how much to set aside in reserve for a claim that has been filed but not decided, but that adjuster may not indicate the express reasoning behind the amount reserved. A high reserve equally indicates a large probability of winning a small amount or a small probability of winning a very large amount, whereas only the former case may have merit by our definition (if $p_c > 0.50$).

Fourth, even when they do express judgments of claim merit, insurers, attorneys, and others typically do so using categorical variables that take only a few values. For example, Phillips Jr. et al. (2004, 122) relied on pooled medical malpractice claims data from multiple insurers in which an insurer's physicians had reviewed the file and judged whether there had been medical negligence and whether the claim was "defensible" or "indefensible." Similarly, Taragin et al. (1992, 780–81) relied on judgments by physicians and claims adjusters in the claims files of a physician-owned insurance company to classify the claim as "indefensible—insured admits deviation," "no peer review—clearly defensible," "defensible," "indefensible," or "defensibility unclear."

Fifth, there is the task of establishing whether these contemporaneous estimates are valid and reliable. For example, Ross (1980, 98–101) observed how, rather than undertaking the complex inquiry required by the negligence law on the books, the auto insurance claims adjusters he interviewed routinely assumed that in rear-end collision claims, the driver of the striking vehicle is liable for negligence. To be sure, this rule of thumb may validly predict what would happen if those rear-end collision cases were decided in court on the merits. On the other hand, to the extent that claims adjuster evaluations depend on factors other than the merits (such as the cyclical behavior of the insurance market), this approach will be subject to bias. Even lawyer predictions of litigation outcomes are subject to overconfidence and optimism bias (Goodman-Delahanty et al. 2010). Moreover, based on four case studies of medical malpractice claims that went to trial, Vidmar (1995, 67) suggests that "there is often a great deal of uncertainty in the process [of litigating medical malpractice cases]," so that attorney

"estimates of case strength may reverse . . . at any time as . . . more information is uncovered."

As an example of the problems confronting researchers who rely on estimates of claim merits made by situated actors, consider the Equal Employment Opportunity Commission (EEOC), a federal administrative agency. Most federal employment discrimination laws require filing discrimination charges with the EEOC as a precondition for bringing a lawsuit, and mandate the EEOC to identify whether filed charges of discrimination have "reasonable cause." Pursuant to that mandate, the EEOC, since 1995, has assigned priority to each new charge filed by coding it A, B, or C. Category A covers "charges that fall within the [EEOC's] national or local enforcement plan," charges "where further investigation will probably result in a cause finding," or charges for which "irreparable harm will result unless processing is expedited." Category C covers charges for which the agency "has sufficient information from which to conclude that it is not likely that further investigation will result in a cause finding." Category B covers charges that "initially appear to have some merit but will require additional evidence to determine whether continued investigation is likely to result in a cause finding." After further investigation, the EEOC instructs agency staff to recode these charges as either A or C (Equal Employment Opportunity Commission 1995, 405:7313).

Ullman et al. found that for all Americans with Disabilities Act charges filed with the EEOC between June 1995 and March 1998, charges involving anxiety, depression, and manic-depression were less likely to be coded as Category A than those that did not involve those issues (2001, 648). The authors suggested several explanations for this finding. One was that there were real differences of merit between the ADA cases that did involve these psychiatric issues and those that did not. Another was that "stereotypes or stigma" about psychiatric conditions influenced agency staff, merits notwithstanding (648). Still another was that claimants with these psychiatric conditions are less able to clearly articulate their discrimination allegations in the charges they file, which affects the way agency staff members assign preinvestigation priority codes (648–49).

Testing these explanations requires validating whether EEOC staff assign priority codes in accordance with the available information on claim merit. One way to test these explanations would be to take a sample of EEOC charge files, have independent expert raters (e.g., employment lawyers) code them using the EEOC priority codes, and test for agreement with the previously assigned EEOC code for each file.

Another approach might use hypothetical vignettes to capture the set of circumstances that influence the way EEOC staff assign priority codes. In vignette studies, researchers present respondents with a short hypothetical scenario and ask them to judge whether the law permits or prohibits the conduct described therein (York 1989; Schmedemann and Parks 1994; Kim 1997, 1999). This method has long been used in social research (Herskovits 1950). In a more sophisticated variant, the factorial survey, researchers systematically vary some aspects of the facts presented in each vignette, allowing them to discover which facts influence respondents' judgments (Jasso 2006; Wallander 2009).

Still another approach is to match EEOC charges with subsequent litigation, and compare the assigned EEOC priority codes with subsequent litigation outcomes. For example, from a random sample of employment civil rights cases filed in federal court

between 1988 and 2003, Nielsen, Nelson, and Lancaster (2010, 191) matched the EEOC priority codes for those cases with case outcomes, and concluded, based on a discrete hazard model, that the priority codes “had little explanatory effect” on the probability that a case would end because of dismissal, early settlement (prior to motion for summary judgment), plaintiff loss at summary judgment, or post-summary-judgment settlement. In particular, they found no statistically significant effect of EEOC priority codes A and C on the probability of each of those case outcomes (186). “The only statistically significant effect that EEOC priority codes have is that B cases are more likely to obtain early settlement than other cases” (191).⁴

Administrative agency estimates also raise the particular problem that agency staff may prioritize complaints based on considerations other than the apparent merit of the complaint. Some agencies select complaints for investigation and litigation to advance an agenda to move “the law” in a particular direction. For example, EEOC priority code A applies not only to charges that are likely to result in a “reasonable cause” finding, but also to charges that fall within the EEOC’s national enforcement plan. That plan identifies as enforcement priorities cases that may “promote the development of law supporting the antidiscrimination purposes of the statutes” that the agency enforces, such as claims presenting certain “unresolved issues of statutory interpretation” under one or more of those statutes (Equal Employment Opportunity Commission 1996, 405:7354). In other words, the agency may decide to litigate a case to persuade a court or other decisionmaker to change the underlying legal decision rules under which its odds of success are currently poor or uncertain. In sum, the agency’s ranking may well reflect in part its own priorities, rather than being simply a guess about the charge’s actual merits under current law.

C. Inference from Observable Claim Characteristics

This strategy relies on a close correlation between claim merit and easily observable or self-reported characteristics of potential and actual claims. To illustrate this strategy, we consider research on (1) the “take-up” of social benefits, (2) sexual harassment, and (3) insurance fraud.

1. *The “Take-Up” Rate of Social Benefits*

Researchers such as J. Currie (2006), Orbach (2006), Remler and Glied (2003), and Craig (1991) describe the extent to which people do not apply for a benefit (e.g., unemployment insurance, Medicaid) for which they are eligible by calculating a “take-up” rate for that benefit. A take-up rate requires a numerator—typically the number of

4. It may be that the EEOC codes are valid and reliable estimates of claim merit, given the information available at the time the EEOC staff assigned those codes. During subsequent litigation, lawyer investment in fact finding and civil discovery may increase information about a case’s merits. If so, we should expect substantial differences between estimates of a claim’s odds of success based only on information in the EEOC charge file and such estimates at different stages of actual litigation.

people who actually received the benefit—and a denominator—the total number of people who are eligible to receive the benefit. For that denominator, since legal rules set the conditions for eligibility, researchers need data indicating the presence or absence of the eligibility conditions when they sample for benefit-eligible persons (or classify persons in an already drawn sample as benefit eligible).

These studies assume that by applying the algorithm defined by the eligibility rules, the researchers can predict the outcome of any individual application for benefits. The benefit-granting institution would more likely than not grant benefits ($p_c > 0.50$) for those who meet the eligibility formula, and is more likely than not to deny benefits ($p_c \leq 0.50$) for those who do not. Such studies often go on to model claiming processes, and to estimate the degree to which such variables as stigma, application process complexity, and benefit size affect benefit take-up (Kerr 1982; Moffitt 1983; Duclos 1995). In general, these studies conclude that take-up rates are substantially less than 100 percent among those eligible, which constitutes evidence of significant underclaiming.

As benefit eligibility criteria become more complex, researchers must work harder to provide an accurate specification of the benefit-eligible class, especially if eligibility criteria diverge from self-reported characteristics in household survey data. Examples of more complex eligibility rules include those governing Social Security Disability Insurance (SSDI) benefits (Dwyer et al. 2003), “medical need” under Medicaid (Ettner 1997), and workers’ compensation benefits (Shannon and Lowe 2002).

In some studies, researchers combine survey self-reports with administrative agency data on filing and benefit decisions. This approach permits some estimate of the bias of survey self-reports as eligibility measures. For example, Benítez-Silva et al. studied, from a large panel survey of older individuals about their health and retirement, a subsample of individuals who self-reported disability status and applied for SSDI benefits or Supplemental Security Income (SSI) benefits (2004, 650, 654). Comparing the survey self-reports of disability status with the way the Social Security Administration (SSA) actually decided the respondents’ benefit applications, they found that roughly 15 percent of the people in their subsample reported no disability but still collected benefits, while roughly 18 percent self-reported that they were disabled but were denied benefits (651, Table 1). Nevertheless, Benítez-Silva et al. concluded that the disability self-reports were an unbiased estimator of the ultimate SSA eligibility decision because “overall, the individual’s assessment of their own health condition matches that of the SSA” (650). However, since this study did not randomly sample SSDI or SSI applicants, its subsample may not be representative of all such applicants.

The take-up rate literature does not typically estimate overclaims, and therefore its findings do not rule out at least some overclaiming of benefits. For example, based on data from a computerized system that identified discrepancies between AFDC and Food Stamp recipients’ self-reported earnings and employer-reported wage records, which in turn triggered fraud investigations, Wolf and Greenberg (1986, 452) estimated that only 2 to 3 percent of the average combined Food Stamp and AFDC caseloads in two New Jersey counties between January 1981 and March 1981 involved fraudulent underreporting of earnings. Another example: In randomized trials involving unemployment insurance (UI) claimants at sites in four states (Connecticut, Massachusetts, Virginia, and Tennessee), Ashenfelter, Ashmore, and Deschênes randomly subjected UI claimants to “a number of additional verifications of initial and continuing eligibility” before

the relevant state agency issued the first UI payment to them (2005, 56). This resulted in about a 3 percent overall reduction in the overall benefit qualification rate for the claimants, “mainly reflecting [a] large 8% reduction in Connecticut” (67), which suggests some limited UI overclaiming.

2. Sexual Harassment

Sexual harassment research often asks survey respondents to rate their experiences in terms of the severity of the harassment, which ranges from verbal statements with sexually explicit content to actual physical assault (Antecol and Cobb-Clark 2004; Blackstone, Uggen, and McLaughlin 2009). To hold an employer liable for workplace sexual harassment under Title VII of the 1964 Civil Rights Act, an employee must show employer-attributable conduct that is “severe or pervasive” enough that “a reasonable person would find [it] hostile or abusive,” as determined by factors including the frequency and severity of underlying conduct; “whether it is physically threatening or humiliating, or a mere offensive utterance; and whether it unreasonably interferes with an employee’s work performance” (*Harris v. Forklift Systems* 1993).

Accordingly, the more sensitive the survey instrument is to the severity of the respondents’ self-reported experiences, the better one can predict respondents’ odds of success on any Title VII sexual harassment claim they might bring. Where studies collect data on respondent responses to the alleged harassing behavior (from doing nothing to complaining to a supervisor to filing a lawsuit), they also provide a rough starting point for estimating the number of sexual harassment underclaims.

To move past this starting point, however, the studies must have also collected other data relevant to the way the claims-processing institution in question would decide the claim. For example, suppose a survey respondent clearly suffered an experience that would satisfy the underlying decision rule about what constitutes sexual harassment under Title VII. Even if the respondent did not complain about the conduct in question, however, her potential Title VII suit might nevertheless not count as an underclaim ($p_c < 0.50$) because Title VII affords a defense against some kinds of sexual harassment liability if the employer provided, but the aggrieved worker did not pursue, an internal grievance procedure for responding to sexual harassment (*Faragher v. City of Boca Raton* 1998).

3. Insurance Fraud

Research on insurance fraud amounts to estimating the number of overclaims₂ processed by a particular kind of institution, namely, insurers’ claims departments. Such estimates are based on observing the characteristics of insurance claims and comparing them to those of verified instances of insurance fraud.

Verified instances of insurance fraud are easy to find. For example, in *In re Silica Products Liability Litigation* (S.D. Tex. 2005), a district court found that 111 claims of silicosis—involving roughly 10,000 individuals—were based on diagnoses that were “manufactured for money,” pursuant to a scheme in which “lawyers, doctors, and

screening companies were all willing participants" (635). For another example, Dornstein (1998, ch. 6) discusses the "Friends of the Friendless," a large and sophisticated fraudulent auto accident enterprise in southern California, involving several doctors and lawyers.

It is much harder, however, to find out how often insurance fraud occurs. Imagine asking a survey respondent, "Did you ever file a claim for an injury that you did not suffer?" or "Have you ever exaggerated the severity of an injury you did experience?" Answers to such questions would likely be unreliable. Indeed, since intentional overclaimers presumably try to avoid detection, prevalence estimates based on detected overclaims₂ are likely to be biased downward. Reviewing the social science literature on insurance fraud, Tennyson concludes that although "conventional wisdom often estimates the prevalence of insurance claims fraud at about 10 percent of claims, . . . this statistic appears to be more folk wisdom than fact" (2008, 1184).

In a sophisticated study, Caron and Dionne (1999) estimate how often auto insurance fraud occurs based on insurer data on suspected fraud in a sample of 2,509 randomly investigated claims files in the Quebec auto insurance industry. Using an ingenious statistical technique that allows for simultaneous estimation of the detection rate and the underlying fraud or overclaiming₂ rate, they estimate that roughly 12 percent of all claims involve some kind of fraud, of which insurers detect only about one-third. If they generalize, these results suggest that even trained investigators catch only a small proportion of overclaims₂, so existing estimates may indeed have a downwards bias.

Insurance companies often attempt to detect fraudulent auto insurance claims (overclaims₂) by generating suspicion scores based on claim characteristics that have been closely correlated with insurance fraud in past cases. For example, Dornstein quotes insurance industry trade literature from 1979 that offer a rough list of slip-and-fall fraud indicators: "Claimant is a transient or out-of-towner on vacation"; "Claimant is overly pushy and demanding for a quick settlement"; "Typical defrauder will claim that as a result of his fall, he suffered personal injury, lost time from work, and broke his eyeglasses, dentures, or other expensive props, such as cameras or wristwatches" (1998, 267).

More sophisticated versions of this kind of checklist are based on mathematical techniques that combine information on attributes of a claim that are statistically associated with suspected fraud (Weisberg and Derrig 1991). For example, Brockett, Xia, and Derrig (1998) use sixty-five indicator ("yes/no") variables to predict the presence of fraud by means of an artificial intelligence technique based on neural networks. Insurers rarely, if ever, decisively verify the existence of fraud by bringing suspected claims before a court for an official determination that fraud has occurred (but see Artís, Ayuso, and Guillén 2002). That would require extensive investigation and legal procedures, both of which are extremely expensive. Instead, they classify claims as "suspicious" based on the ratings of experienced claims adjusters. They then use statistical techniques to discover which attributes of a claim predict the adjusters' suspicion of fraud, rather than predicting actual (verified) instances of fraud itself. If this is indeed their approach, insurers' fraud detection methods are ultimately no better at capturing overclaiming than self-report survey studies are as measures of underclaiming. In neither case are merit estimates verified by independent investigation.

D. Field Experiments or Audit Studies

Field experiments (such as audits) are one of the best ways to measure underclaims. To illustrate, under laws prohibiting discrimination in labor, housing, and product markets, the merit of a claim of discrimination often turns on the strength of the inference that someone was treated worse than they would have been had they been of a different race or sex. Since this counterfactual cannot be observed directly, job applicants or those seeking to rent an apartment often find it impossible to detect the presence of discrimination. Someone who is turned down for a job will typically lack any information about the employer's hiring decision, and will thus be unable to infer why he or she was not hired. Survey or interview methods face the same problem. Asking survey respondents whether they have experienced discrimination is unlikely to produce accurate answers when the discrimination is difficult to detect.

For over forty years, some researchers have turned to field experiments to try to surmount this problem (Raich and Rich 2002). They send out pairs of individuals to apply for jobs or mortgages, or seek to buy homes, cars, or cups of coffee, among other things. Then, they compare how the pairs were treated. If A and B are "identical" except for race; both apply for the same jobs; and A is systematically offered a job more often than B is, this is a strong basis to infer illegal race discrimination against B. To be sure, at any particular employer, A might randomly receive more favorable treatment than B, or vice-versa. Thus, audit studies, if properly implemented, provide good evidence of aggregate (net) discrimination against one or the other group, although without conclusively demonstrating discrimination in any particular encounter (Heckman and Siegelman 1993, 187).

Audit studies of this kind have been used to detect discrimination in many situations, including employment (Bertrand and Mullainathan 2004; Pager, Bonikowski, and Western 2009; Carlsson 2010), housing (Ondrich, Ross, and Yinger 2000), auto sales (Ayres and Siegelman 1995), and waiting times at a retail coffee seller (Myers et al. 2008), as well as tipping by taxicab riders (Ayres, Vars, and Zakariya 2005) and by restaurant customers (Lynn et al. 2008). They typically show that in the aggregate, discrimination occurs often in all these contexts. Yet, virtually no anti-discrimination lawsuits are filed against discriminatory tipping practices. Hiring discrimination is also litigated relatively rarely (Nielsen, Nelson, and Lancaster 2010), as is housing discrimination. In these contexts, therefore, there are likely to be claims with merit that are never brought, making a persuasive case for at least some underclaiming. Of course, it may be that even though nobody brings meritorious hiring discrimination claims, those hiring discrimination claims that *are* brought nevertheless lack merit. The audit evidence does not rule out some overclaiming, although *net* overclaiming seems highly unlikely.

Field experiments can detect other kinds of illegal behavior as well. For example, by sending cars in need of specific predetermined repairs to various auto mechanics, researchers can detect the incidence of unnecessary repairs as well as incidence of undiagnosed (but needed) repairs. Using this strategy, Schneider (forthcoming) found that roughly one-third of all mechanics recommended unnecessary repairs. A similar strategy uncovered substantial fraud in television repair (Phelan 1974). If almost no

one files claims based on improperly performed repairs, we can infer substantial underclaiming.

E. Inference from the Characteristics of Filed Claims

Still another strategy is to infer the merits of filed but undecided claims from the characteristics of filed claims that were settled or decided on the merits.⁵ For example, it is reasonable to assume that, all else equal, claims with a greater probability of success on the merits will command a higher settlement value than those with a lower probability of success. Eisenberg and Lanvers review evidence of the relationship between case quality and settlement outcome and conclude that “the best available evidence is that within-category settlement rates are highly sensitive to the merits of the case” (2009, 124). If so, the characteristics and money values of settled claims can provide some insight into how such claims might have fared had they instead been decided on the merits. However, since the probability of success for different claims varies widely and claims are not randomly selected for settlement or decision, this empirical strategy requires a set of assumptions—a model—of the processes by which claims are settled or pursued.

The economic literature assumes that potential legal claims are selected for litigation based largely on their underlying merits as perceived by the parties. Given a model of the mechanisms by which cases are selected for litigation—and there is no consensus about the appropriate model (Bebchuk 1984; Shavell 1996; Waldfogel 1998)—it is possible, using knowledge of outcomes in decided claims, to estimate the merits of claims that are *filed* but not decided.

For example, Siegelman and Waldfogel (1999) use data on groups of cases filed in one federal district court to test a version of the model proposed by Priest and Klein (1984). That model predicts both the plaintiff win rate in tried cases, and the rate at which cases go to trial, as functions of three primitives: (1) the fraction of filed cases that would yield pro-plaintiff verdicts if tried (the decision standard), (2) the ability of the parties to estimate the quality of their own cases (uncertainty), and (3) the degree of stakes asymmetry (the relative amount that each side stands to win or lose by virtue of the litigation) between parties. Siegelman and Waldfogel test this model by collecting data on six distinct types of disputes (e.g., contracts cases) and characterizing each according to plausible proxies for the primitives in the Priest/Klein model. (For example, the degree of stakes asymmetry is measured by the share of institutional [nonindividual] parties among all plaintiffs and defendants, and by the proportion of repeat players among these groups.) They then empirically estimate a structural version of the Priest/Klein model, and compare the estimated parameters with their independently gathered evidence. In general, the model tracks the independent evidence relatively well. The authors conclude that if all filed cases were actually tried, plaintiffs would win 0.8 percent of prisoner cases, 8.9 percent of civil rights cases, and 43.7 percent of contracts cases (Siegelman and Waldfogel 1999, Table 6). Other studies in this vein include Kessler, Meites, and Miller (1996) and Eisenberg and Farber (1997).

5. We are grateful to a referee for suggesting this discussion.

This kind of analysis has several limitations. First, it only helps estimate the merits of filed cases, without having much—if anything—to say about claims that are never brought (potential claims). Second, these estimates depend heavily on the model's assumptions about the selection process, and in particular about the information available to parties and the process of settlement bargaining. These assumptions are not directly testable.

Moreover, this approach has relatively little to say about the merits of potential claims at lower levels of the dispute pyramid. Thus, suppose that if all filed prisoner cases went to trial, plaintiffs would prevail in only 0.8 percent of them, as Siegelman and Waldfogel (1999) estimate. Imagine a world in which 2,000 prisoners suffer sexual abuse, but never file a claim because of fear, shame, or lack of resources. An additional 1,000 prisoners have complaints about food or conditions of incarceration, and of this group, all file a claim. Two-hundred of these claims are formally adjudicated, and the plaintiff prevails in two of them (1 percent). Even if the success rate in filed-but-not-adjudicated claims is only 0.8 percent, this says nothing about the merits of the 2,000 potential claims that are never filed, which come from a completely different population of disputes. Without sociolegal knowledge of the selection process, therefore, it is impossible to make reliable inferences about the merits of unbrought claims based only on information about decided cases.

IV. CONCLUSION

Underclaiming or overclaiming arguments assert that courts or other claims-processing institutions see too few claims with merit or too many claims without merit. This article has identified the conceptual structure of these arguments and canvassed ways of estimating the merits of potential and actual claims.

In so doing, this article has also made it easier to separate valid underclaiming and overclaiming arguments from ones that are largely stalking horses for or against the purpose, scope, or morality of a claims-processing institution's rule of decision. Underclaiming and overclaiming arguments have rhetorical power precisely because they attempt to avoid debates about the propriety of an institution's decision rules, focusing instead on the quality of claims brought under those rules. For example, it is much easier for opponents of tort or civil rights laws to argue that the laws permit too many meritless cases to be brought than to argue against the fundamental purpose of such laws. To be sure, when assigning social values to underclaims and overclaims, valid underclaiming and overclaiming arguments must ultimately face normative disputes about the purpose, scope, and morality of a claims-processing institution's decision rules, and the tradeoffs they may embody. Yet, by pursuing valid estimates of claim merit, we believe that sociolegal research can substantially increase what we know, and more sharply identify what we do not know, about how well claims-processing institutions encourage claims with merit and discourage those without it.

REFERENCES

Abel, Richard L. 1987. The Real Tort Crisis—Too Few Claims. *Ohio State Law Journal* 48(2): 443–67.

- Amoros, Emmanuelle, Jean-Louis Martin, Mireille Chiron, and Bernard Laumon. 2007. Road Crash Casualties: Characteristics of Police Injury Severity Misclassification. *Journal of Trauma: Injury, Infection, and Critical Care* 62(2): 482–90.
- Antecol, Heather, and Deborah Cobb-Clark. 2004. The Changing Nature of Employment-Related Sexual Harassment: Evidence from the US Federal Government, 1978–1994. *Industrial and Labor Relations Review* 57(3): 443–61.
- Artís, Manuel, Mercedes Ayuso, and Montserrat Guillén. 2002. Detection of Automobile Insurance Fraud with Discrete Choice Models and Misclassified Claims. *Journal of Risk and Insurance* 69(3): 325–40.
- Ashenfelter, Orley, David Ashmore, and Olivier Deschênes. 2005. Do Unemployment Insurance Recipients Actively Seek Work? Evidence from Randomized Trials in Four U.S. States. *Journal of Econometrics* 125(1–2): 53–75.
- Ayres, Ian, and Peter Siegelman. 1995. Race and Gender Discrimination in Bargaining for a New Car. *American Economic Review* 85(3): 304–21.
- Ayres, Ian, Fredrick E. Vars, and Nassar Zakariya. 2005. To Insure Prejudice: Racial Disparities in Taxicab Tipping. *Yale Law Journal* 114(7): 1613–76.
- Baker, Tom. 2005. Reconsidering the Harvard Medical Practice Study About the Validity of Medical Malpractice Claims. *Journal of Law, Medicine, and Ethics* 33(3): 501–14.
- Beck, Allen J., and Paige M. Harrison. 2010. *Sexual Victimization in Prisons and Jails Reported by Inmates, 2008–09*. Washington, DC: U.S. Department of Justice.
- Benítez-Silva, Hugo, Moshe Buchinsky, Hiu M. Chan, Sofia Cheidvasser, and John Rust. 2004. How Large Is the Bias in Self-Reported Disability? *Journal of Applied Econometrics* 19(6): 649–70.
- Berg, Marc, and Paul Haterink. 2004. Embodying the Patient: Records and Bodies in Early 20th-Century US Medical Practice. *Body and Society* 10(2–3): 13–41.
- Bertrand, Marianne, and Sendhil Mullainathan. 2004. Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *American Economic Review* 94(4): 991–1013.
- Blackstone, Amy, Christopher Uggen, and Heather McLaughlin. 2009. Legal Consciousness and Responses to Sexual Harassment. *Law & Society Review* 43(3): 631–68.
- Brake, Deborah L., and Joanna L. Grossman. 2008. The Failure of Title VII as a Rights-Claiming System. *North Carolina Law Review* 86(4): 859–935.
- Brennan, Troyen A., Lucian L. Leape, Nan M. Laird, Liesi Hebert, A. R. Localio, Ann G. Lawthers, Joseph P. Newhouse, Paul C. Weiler, and Howard H. Hiatt. 1991. Incidence of Adverse Events and Negligence in Hospitalized Patients. *New England Journal of Medicine* 324(6): 370–76.
- Brennan, Troyen A., Colin M. Sox, and Helen R. Burstin. 1996. Relation Between Negligent Adverse Events and the Outcomes of Medical-Malpractice Litigation. *New England Journal of Medicine* 335(26): 1963–67.
- Brickman, Lester. 2011. *Lawyer Barons: What Their Contingency Fees Really Cost America*. New York: Cambridge University Press.
- Brockett, Patrick L., Xiaohua Xia, and Richard A. Derrig. 1998. Using Kohonen's Self-Organizing Feature Map to Uncover Automobile Bodily Injury Claims Fraud. *Journal of Risk and Insurance* 65(2): 245–74.
- Carlsson, Magnus. 2010. Experimental Evidence of Discrimination in the Hiring of First- and Second-Generation Immigrants. *LABOUR* 24(3): 263–78.
- Caron, Louis, and Georges Dionne. 1999. Insurance Fraud Estimation: More Evidence from Quebec Automobile Insurance Industry. In *Automobile Insurance: Road Safety, New Drivers, Risks, Insurance Fraud and Regulation*, eds. Georges Dionne and Claire Laberge-Nadeau, 175–82. Norwell, MA: Kluwer Academic Publishers.
- Cheney, Frederick W., Karen Posner, Robert A. Caplan, and Richard J. Ward. 1989. Standard of Care and Anesthesia Liability. *JAMA: The Journal of the American Medical Association* 261(11): 1599–1603.
- Colvin, Alexander J. 2011. An Empirical Study of Employment Arbitration: Case Outcomes and Processes. *Journal of Empirical Legal Studies* 8(1): 1–23.

- Craig, Barbara L. 1989–1990. Hospital Records and Record-Keeping, c. 1850–c. 1950, Part 1: The Development of Records in Hospitals. *Archivaria* 29:57–87.
- . 1990. Hospital Records and Record-Keeping, c. 1850–c. 1950, Part II: The Development of Record-Keeping in Hospitals. *Archivaria* 30:21–38.
- Craig, Peter. 1991. Costs and Benefits: A Review of Research on Take-Up of Income-Related Benefits. *Journal of Social Policy* 20(4): 537–65.
- Cranberg, Lee D., Thomas H. Glick, and Luke Sato. 2007. Do the Claims Hold Up? A Study of Medical Negligence Claims against Neurologists. *Journal of Empirical Legal Studies* 4(1): 155–62.
- Currie, Ab. 2006. A National Survey of the Civil Justice Problems of Low- and Moderate-Income Canadians: Incidence and Patterns. *International Journal of the Legal Profession* 13(3): 217–42.
- Currie, Janet. 2006. The Take-Up of Social Benefits. In *Public Policy and the Income Distribution*, eds. Alan J. Auerbach, David Card, and John M. Quigley, 80–148. New York: Russell Sage Foundation.
- Dornstein, Ken. 1998. *Accidentally, on Purpose: The Making of a Personal Injury Underworld in America*. New York: St. Martin's Press.
- Duclos, Jean-Yves. 1995. Modelling the Take-Up of State Support. *Journal of Public Economics* 58(3): 391–415.
- Dwyer, Debra, Jianting Hu, Denton R. Vaughan, and Bernard Wixon. 2003. Counting the Disabled: Using Survey Self-Reports to Estimate Medical Eligibility for Social Security's Disability Programs. *Journal of Economic and Social Measurement* 28(3): 109–42.
- Eisenberg, Theodore, and Henry Farber. 1997. The Litigious Plaintiff Hypothesis: Case Selection and Resolution. *Rand Journal of Economics* 28:S92–S112.
- Eisenberg, Theodore, and Charlotte Lanvers. 2009. What Is the Settlement Rate and Why Should We Care? *Journal of Empirical Legal Studies* 6(1): 111–46.
- Equal Employment Opportunity Commission. 1995. Priority Charge Handling Procedures. In *Fair Employment Practices Manual*, 405:7311–405:7322. Washington, DC: Bureau of National Affairs, Inc.
- . 1996. National Enforcement Plan. In *Fair Employment Practices Manual*, 405:7351–405:7358. Washington, DC: Bureau of National Affairs, Inc.
- Ettner, Susan L. 1997. Medicaid Participation Among the Eligible Elderly. *Journal of Policy Analysis and Management* 16(2): 237–55.
- Felstiner, William L., Richard L. Abel, and Austin Sarat. 1980. The Emergence and Transformation of Disputes: Naming, Blaming, Claiming. . . . *Law & Society Review* 15(3/4): 631–54.
- Gandhi, Tejal K., Allen Kachalia, Eric J. Thomas, Ann L. Puopolo, Catherine Yoon, Troyen A. Brennan, and David M. Studdert. 2006. Missed and Delayed Diagnoses in the Ambulatory Setting: A Study of Closed Malpractice Claims. *Annals of Internal Medicine* 145(7): 488–96.
- Genn, Hazel. 1999. *Paths to Justice: What Do People Do and Think About Going to Law*. Oxford, UK: Hart Publishing.
- Goodman-Delahunty, Jane, Pär A. Granhag, Maria Hartwig, and Elizabeth F. Loftus. 2010. Insightful or Wishful: Lawyers' Ability to Predict Case Outcomes. *Psychology, Public Policy, and Law* 16(2): 133–57.
- Harris, Donald, Mavis Maclean, Hazel Glenn, Sally Lloyd-Bostock, Paul Fenn, Peter Corfield, and Yvonne Brittan. 1984. *Compensation and Support for Illness and Injury*. Oxford, UK: Clarendon Press.
- Harvard Medical Practice Study. 1990. *Patients, Doctors, and Lawyers: Medical Injury, Malpractice Litigation, and Patient Compensation in New York: The Report of the Harvard Medical Practice Study to the State of New York*. New York: Harvard Medical Practice Study.
- Heckman, James J., and Peter Siegelman. 1993. The Urban Institute Audit Studies: Their Methods and Findings. In *Clear and Convincing Evidence: Measurement of Discrimination in America*, eds. Michael Fix and Raymond Struyk. Washington, DC: Urban Institute.
- Herskovits, Melville J. 1950. The Hypothetical Situation: A Technique of Field Research. *Southwestern Journal of Anthropology* 6(1): 32–40.
- Jasso, Guillermina. 2006. Factorial Survey Methods for Studying Beliefs and Judgments. *Sociological Methods & Research* 34(3): 334–423.

- Kerr, Scott A. 1982. Predicting Welfare Benefit Claiming Using Expectancy Theory. *Journal of Economic Psychology* 2(4):287–97.
- Kessler, Daniel, Thomas Meites, and Geoffrey Miller. 1996. Explaining Deviations from the Fifty-Percent Rule: A Multimodal Approach to the Selection of Cases for Litigation. *Journal of Legal Studies* 25(1): 233–59.
- Kim, Pauline T. 1997. Bargaining with Imperfect Information: A Study of Worker Perceptions of Legal Protection in an At-Will World. *Cornell Law Review* 83(1): 105–55.
- . 1999. Norms, Learning, and Law: Exploring the Influences on Workers' Legal Knowledge. *University of Illinois Law Review* 1999(2): 447–515.
- Kritzer, Herbert M. 2011. The Antecedents of Disputes: Complaints and Claiming. *Oñati Socio-Legal Series* 1(6): 1–31.
- Legal Services Corporation. 2009. *Documenting the Justice Gap: The Current Unmet Civil Legal Needs of Low-Income Americans*. Washington, DC: Legal Services Corporation.
- Localio, A. R., Ann G. Lawthers, Troyen A. Brennan, Nan M. Laird, Liesi E. Hebert, Lynn M. Peterson, Joseph P. Newhouse, Paul C. Weiler, and Howard H. Hiatt. 1991. Relation Between Malpractice Claims and Adverse Events Due to Negligence. *New England Journal of Medicine* 325(4): 245–51.
- Lynn, Michael, Michael Sturman, Christie Ganley, Elizabeth Adams, Matthew Douglas, and Jessica McNeil. 2008. Consumer Racial Discrimination in Tipping: A Replication and Extension. *Journal of Applied Social Psychology* 38(4): 1045–60.
- Miller, Richard E., and Austin Sarat. 1980. Grievances, Claims, and Disputes: Assessing the Adversary Culture. *Law & Society Review* 15(3): 525–66.
- Moffitt, Robert. 1983. An Economic Model of Welfare Stigma. *American Economic Review* 73(5): 1023–35.
- Moss, Phillip, and Chris Tilly 2006. Learning About Discrimination by Talking to Employers. In *Handbook on the Economics of Discrimination*, eds. William M. Rodgers, 61–96. Cheltenham, UK: Edward Elgar Publishing.
- Murayama, Masayuki. 2007. Experiences of Problems and Disputing Behavior in Japan. *Meiji Law Journal* 14:1–59.
- Myers, Caitlin K., Marcus Bellows, Hiba Fakhoury, Douglas Hale, Alexander Hall, and Kaitlin Ofman. 2008. Ladies First? A Field Study of Discrimination in Coffee Shops. *Applied Economics* 42(14): 1–9.
- Newman, Jon O. 1996. Pro Se Prisoner Litigation: Looking for Needles in Haystacks. *Brooklyn Law Review* 62(2): 519–27.
- Nielsen, Laura B., Robert L. Nelson, and Ryon Lancaster. 2010. Individual Justice or Collective Legal Mobilization?: Employment Discrimination in the Post Civil Rights United States. *Journal of Empirical Legal Studies* 7(2): 175–201.
- Ondrich, Jan, Stephen L. Ross, and John Yinger. 2000. How Common Is Housing Discrimination? Improving on Traditional Measures. *Journal of Urban Economics* 47(3): 470–500.
- Orbach, Barack Y. 2006. Unwelcome Benefits: Why Welfare Beneficiaries Reject Government Aid. *Law and Inequality: A Journal of Theory and Practice* 24(1): 107–55.
- Pager, Devah, Bart Bonikowski, and Bruce Western. 2009. Discrimination in a Low-Wage Labor Market. *American Sociological Review* 74(5): 779–99.
- Phelan, John J. 1974. *Economic Report: Regulation of the Television Repair Industry in Louisiana and California: A Case Study: Staff Report to the Federal Trade Commission*. Washington, DC: U.S. Government Printing Office.
- Phillips, R. L. Jr., L. A. Bartholomew, S. M. Dovey, G. E. Fryer Jr., T. J. Miyoshi, and L. A. Green. 2004. Learning from Malpractice Claims About Negligent, Adverse Events in Primary Care in the United States. *Quality and Safety in Health Care* 13(2): 121–26.
- Priest, George L. 1987. The Current Insurance Crisis and Modern Tort Law. *Yale Law Journal* 96(7): 1521–90.
- Priest, George L., and Benjamin Klein. 1984. The Selection of Disputes for Litigation. *Journal of Legal Studies* 13(1): 1–55.
- Raich, P. A., and J. Rich. 2002. Field Experiments of Discrimination in the Marketplace. *Economic Journal* 112(483): F480–F515.

- Remler, Dahlia K., and Sherry A. Glied. 2003. What Other Programs Can Teach Us: Increasing Participation in Health Insurance Programs. *American Journal of Public Health* 93(1): 67–74.
- Roehling, Mark V. 2002. The Good Cause Norm in Employment Relations: Empirical Evidence and Policy Implications. *Employee Responsibilities and Rights Journal* 14(2–3): 91–104.
- Ross, H. L. 1980. *Settled Out of Court: The Social Process of Insurance Claims Adjustment*, 2nd ed. Hawthorne, NY: Aldine Transaction.
- Saks, Michael J. 1992. Do We Really Know Anything About the Behavior of the Tort Litigation System—And Why Not? *University of Pennsylvania Law Review* 140(4): 1147–1289.
- Sandefur, Rebecca L. 2008. Access to Civil Justice and Race, Class, and Gender Inequality. *Annual Review of Sociology* 34(1): 339–58.
- Schmedemann, Deborah A., and Judi M. Parks. 1994. Contract Formation and Employee Handbooks: Legal, Psychological, and Empirical Analyses. *Wake Forest Law Review* 29(3): 647–718.
- Schneider, Henry S. Forthcoming. Agency Problems and Reputation in Expert Services: Evidence from Auto Repair. *Journal of Industrial Economics*.
- Sciortino, Stanley, Mary Vassar, Michael Radetsky, and M. M. Knudson. 2005. San Francisco Pedestrian Injury Surveillance: Mapping, Under-Reporting, and Injury Severity in Police and Hospital Records. *Accident Analysis & Prevention* 37(6): 1102–13.
- Shannon, Harry S., and Graham S. Lowe. 2002. How Many Injured Workers Do Not File Claims for Workers' Compensation Benefits? *American Journal of Industrial Medicine* 42(6): 467–73.
- Shavell, Steven. 1996. Any Frequency of Plaintiff Victory at Trial Is Possible. *Journal of Legal Studies* 25(2): 493–501.
- Siegelman, Peter, and Joel Waldfogel. 1999. Toward a Taxonomy of Disputes: New Evidence Through the Prism of the Priest/Klein Model. *Journal of Legal Studies* 28(1): 101–30.
- Studdert, David M., and Michelle M. Mello. 2007. When Tort Resolutions Are Wrong: Predictors of Discordant Outcomes in Medical Malpractice Litigation. *Journal of Legal Studies* 36(S2): S47–S78.
- Studdert, David M., Michelle M. Mello, Atul A. Gawande, Tejal K. Gandhi, Allen Kachalia, Catherine Yoon, Ann L. Puopolo, and Troyen A. Brennan. 2006. Claims, Errors, and Compensation Payments in Medical Malpractice Litigation. *New England Journal of Medicine* 354(19): 2024–33.
- Studdert, David M., Eric J. Thomas, Helen R. Burstin, Brett I. Zbar, E. J. Orav, and Troyen A. Brennan. 2000. Negligent Care and Malpractice Claiming Behavior in Utah and Colorado. *Medical Care* 38(3): 250–60.
- Taragin, Mark L., Laura R. Willett, Adam P. Wilczek, Richard Trout, and Jeffrey L. Carson. 1992. The Influence of Standard of Care and Severity of Injury on the Resolution of Medical Malpractice Claims. *Annals of Internal Medicine* 117(9): 780–84.
- Tennyson, Sharon. 2008. Moral, Social, and Economic Dimensions of Insurance Claims Fraud. *Social Research: An International Quarterly* 75(4): 1181–1204.
- Thomas, Eric J., Stuart R. Lipsitz, David M. Studdert, and Troyen A. Brennan. 2002. The Reliability of Medical Record Review for Estimating Adverse Event Rates. *Annals of Internal Medicine* 136(11): 812–16.
- Thomas, Eric J., David M. Studdert, Helen R. Burstin, E. J. Orav, Timothy Zeena, Elliott J. Williams, K. M. Howard, Paul C. Weiler, and Troyen A. Brennan. 2000. Incidence and Types of Adverse Events and Negligent Care in Utah and Colorado. *Medical Care* 38(3): 261–71.
- Tsui, K. L., F. L. So, N. N. Sze, S. C. Wong, and T. F. Leung. 2009. Misclassification of Injury Severity Among Road Casualties in Police Reports. *Accident Analysis & Prevention* 41(1): 84–89.
- Ullman, Michael D., Matthew C. Johnsen, Kathryn Moss, and Scott Burris. 2001. The EEOC Charge Priority Policy and Claimants with Psychiatric Disabilities. *Psychiatric Services* 52(5): 644–49.
- Vacco, Dennis C., Frankie Sue Del Papa, Pamela F. Carter, and Christine O. Gregoire. 1995. Letter to the Editor: Free the Courts from Frivolous Prisoner Suits. *New York Times* March 3.
- Vidmar, Neil. 1995. *Medical Malpractice and the American Jury: Confronting the Myths About Jury Incompetence, Deep Pockets and Outrageous Damage Awards*. Ann Arbor: University of Michigan Press.

- Waldfoegel, Joel. 1998. Reconciling Asymmetric Information and Divergent Expectations Theories of Litigation. *Journal of Law and Economics* 41(2): 451–76.
- Wallander, Lisa. 2009. 25 Years of Factorial Surveys in Sociology: A Review. *Social Science Research* 38(3): 505–30.
- Weisberg, Herbert I., and Richard A. Derrig. 1991. Fraud and Automobile Insurance: A Report on the Baseline Study of Bodily Injury Claims in Massachusetts. *Journal of Insurance Regulation* 9(4): 497–541.
- Wolf, Douglas, and David Greenberg. 1986. The Dynamics of Welfare Fraud: An Econometric Duration Model in Discrete Time. *Journal of Human Resources* 21(4): 437–55.
- York, Kenneth M. 1989. Defining Sexual Harassment in Workplaces: A Policy Capturing Approach. *Academy of Management Journal* 32(4): 830–50.
- Zwerling, Craig, and Hillary Silver. 1992. Race and Job Dismissals in a Federal Bureaucracy. *American Sociological Review* 57(5): 651–60.

CASES CITED

- Fragher v. City of Boca Raton*, 524 U.S. 775 (1998).
- Harris v. Forklift Sys.*, 510 U.S. 17 (1993).
- In re Silica Prods. Liab. Litig.*, 398 F. Supp. 2d 563 (S.D. Tex. 2005).
- Shady Grove Orthopedic Assocs., P.A. v. Allstate Ins. Co.*, 130 S. Ct. 1431 (2010).

STATUTES CITED

- 28 U.S.C. § 2072(b) (2006).

APPENDIX

One approach to claim valuation is to let each actual or potential claim's social value vary completely with its probability of success on the merits, sum the resulting values, and thereby generate a single probability-adjusted claiming value. Suppose we survey a total of M individuals, and let:

$$I_j = \begin{cases} 1 & \text{if the } j\text{th person reports experiencing an injury, } j = [1, M]; \\ 0 & \text{otherwise.} \end{cases}$$

$$N = \sum_{j=1}^M I_j$$

$$C_i = \begin{cases} 1 & \text{if the } i\text{th person makes a claim based on their own injury, } j = [1, N]; \\ 0 & \text{otherwise.} \end{cases}$$

$$S_i = \begin{cases} 1 & \text{if the } i\text{th reported claim actually judged to have merit;} \\ 0 & \text{otherwise.} \end{cases}$$

$$F_i = \begin{cases} 1 & \text{if the } i\text{th reported claim actually judged to lack merit;} \\ 0 & \text{otherwise.} \end{cases}$$

$$\sum_{i=1}^N C_i = K = \text{total reported claims}$$

$$\sum_{i=1}^N S_i = T = \text{total reported claims actually judged to have merit}$$

$$\sum_{i=1}^N F_i = V = \text{total reported claims actually judged meritless.}$$

By choosing N , not M , as the upper summation limit for C_i , S_i , and F_i , we ignore those in the sample who did not perceive an injury. Moreover, because some reported claims may be withdrawn before decision or decided on grounds that one may not treat as decisions on the merits (e.g., dismissals on “procedural” grounds), $T + V \leq K$.

To measure for underclaiming (U) and overclaiming (O), we discount each reported injury by its probability of success on the merits (p_j):

$$U = \frac{T}{\sum_{j=1}^M p_j I_j}$$

$$O = \frac{V}{\sum_{j=1}^M (1 - p_j) I_j}.$$

In theory, the final step is to sum those values ($U + O$). This, however, requires agreeing on cost-benefit functions for underclaims and overclaims, as discussed in Section II.B. Absent agreement on how much to value underclaims and overclaims, we cannot have a single probability-weighted claiming value.