The Price of Private Enforcement Under the False Claims Act

August 2012

David Kwok

Abstract

Private plaintiffs and their attorneys are motivated at least in part by monetary awards, but what is the appropriate award for litigation? I consider the case of the False Claims Act, through which the U.S. government recovers losses due to fraud by paying a bounty to whistleblowers. These whistleblowers and their attorneys generally receive a minimum 15% bounty in successful cases. Utilizing a data set of over 3000 fraud cases, I present real world evidence that these whistleblowers are reluctant to bring cases valued under $440,000 and that they would be more likely to do so given a higher bounty percentage. This empirical work suggests that enforcement systems incorporating private initiative should consider scaling percentage bounties to avoid a disproportionate focus on cases with multi-million dollar potential.
The Price of Private Enforcement Under the False Claims Act

I. Introduction: How much should we pay for information?

Combating fraud against the federal government is a difficult task, beginning with the challenge of identifying the occurrence of fraud. “Improper payments” under Medicare and Medicaid alone may start at $70 billion, and these are the more easily detectable cases. Under the False Claims Act (FCA), the Department of Justice (DOJ) has come to rely upon private parties, typically whistleblowers, to identify cases of fraud for prosecution. These whistleblowers, known as “relators” under the statute, receive a percentage of the final recovery against the defendant. The whistleblowers may have a variety of motivations for coming forward with information about these cases of fraud. In this article, I focus on the impact of financial compensation in the whistleblower’s decision. How much should we pay these whistleblowers?

The policy decision to pay whistleblowers has numerous implications. We might be concerned about ensuring sufficient compensation for whistleblowers. We might focus on the societal messages of encouraging disloyalty or rewarding individuals who may be complicit in wrongdoing. For this article, though, I focus narrowly on an instrumental concern of encouraging individuals to blow the whistle: if we pay too much, the government may be leaving “money on the table,” while paying too little might result in much fraud going unreported. Of course, not every potential whistleblower is motivated primarily by money; even without offering rewards, there will be whistleblowers who are sufficiently altruistic or vindictive to come forward with information about fraud.

The empirical challenge for researchers in understanding whistleblower behavior is in experimental design: how to measure the relevance of various factors to the decision to blow the whistle. With regards to the percentage bounty payment, how can we know what amount would be necessary to induce whistleblowing for any particular individual? Some studies utilize survey information, interviewing individual whistleblowers retrospectively. Such studies can provide substantial insight, but they are limited in that they do not address the factors relevant to individuals who did not actually blow the whistle. Other studies apply survey methods to a broader population, trying to determine what individuals say they would do given a hypothetical whistleblower situation. These later methods avoid the problem of only surveying individuals who whistleblow, but it may be difficult to determine whether responses to hypothetical situations correlate with real life action, particularly when the social and career consequences of whistleblowing can be quite severe.

In this study, I utilize an empirical approach based on observed whistleblower behavior that avoids some of the aforementioned challenges. I start with data from the Department

---

2 For a literature review, see Mesmer-Magnus and Viswesvaran, Whistleblowing in Organizations: An Examination of Correlates of Whistleblowing Intentions, Actions, and Retaliation, 62 J. of Business Ethics 277 (2005).
of Justice regarding whistleblower actions under the False Claims Act. Utilizing a change in the tax code in 2004, I look at the change in whistleblower actions in response to the improved post-tax incentives of whistleblowing. From this work, I infer that potential whistleblowers respond to improved post-tax incentives by being more likely to come forward with cases valued under $400,000. To the extent that this response correlates with whistleblowers taking action before fraud becomes worse, the increased effective bounty can result in lower fraud losses for the federal government.

I begin with some background on the False Claims Act.

II. Background on the False Claims Act

The False Claims Act proscribes fraud against the federal government through the imposition of both civil\(^3\) and criminal\(^4\) penalties. Besides traditional public enforcement, the Act also contains *qui tam* provisions, which allow private litigants known as relators to pursue civil actions and prosecute cases of fraud separately from the Department of Justice.\(^5\) Dating back to the Abraham Lincoln presidency, the *qui tam* provisions received renewed attention in 1986 when Congress enhanced the reward structure.\(^6\) Today, relators can receive as much as 30% of the civil recovery, which can be substantial given the treble damages provisions in the statute. Civil penalties also include $5,500 to $11,000 in fines per false claim. Relators do not have to satisfy the traditional requirements of standing;\(^7\) as such, they have a remarkable amount of flexibility in pursuing cases of their choice. From a private enforcement perspective, relators are nearly on par with public enforcement agents in their ability to select cases.\(^8\) The relators, often whistleblowers within an organization, typically obtain representation by counsel on a contingency fee basis; they are not responsible for attorneys' fees if the case is unsuccessful.\(^9\)

The DOJ effectively has a right of first refusal on every FCA *qui tam* case.\(^10\) Upon the initial filing by a relator, the court will immediately seal and stay the case for 60 days. During this time, the DOJ investigates the allegations. The government typically requests time extensions for investigation, which are routinely granted. After an average of 13 months, the DOJ announces whether or not it is "intervening" in the action, also known as its "election" regarding intervention. If it chooses to intervene, it either takes over litigation of the case or dismisses the case outright and may do so over the objections of the relator.\(^11\) If it does not intervene, the relator is then free to litigate the case. Should the relator attempt to settle or dismiss the action, however, she must obtain DOJ consent.\(^12\)

\(^6\) See Beck, 78 N.C.L. REV. 539, 554-65 for a brief history.
\(^8\) Relators are excluded from participating in FCA cases involving tax fraud by statute. See Dennis Ventry, Whistleblowers and Qui Tam for Tax, 61 Tax Lawyer 357 (2008).
\(^10\) See § 3730(b).
\(^11\) § 3730(c)(2)(A).
\(^12\) § 3730(b).
If the DOJ intervenes in the *qui tam* action, the relator is entitled to receive between 15 and 25 percent of the amount recovered.\(^\text{13}\) I refer to this as the “finder's bounty.” If the DOJ declines to intervene in the action and the relator prevails in litigation against the defendant, her share is between 25 to 30 percent.\(^\text{14}\) Regardless of intervention, a successful relator is also entitled to legal fees from the defendant.\(^\text{15}\)

The finder's bounty is the most commonly paid bounty in FCA enforcement; relators typically are unsuccessful in obtaining any recoveries if DOJ declines intervention. The finder's bounty is mostly for the relator's information regarding the fraud as the DOJ runs the litigation. The relator and her attorney will still cooperate with the DOJ after the intervention decision, but I will focus my analysis on the relator's pre-election incentive. A main question for the government, then, is the proper bounty in encouraging relators to come forward with cases of fraud.

### III. A framework for analysis: What is socially optimal?

Before I address the detailed question of the proper bounty, I first consider the broader purpose of the enforcement system. The goal for a hypothetical social planner should be to maximize social welfare. Besides the losses accrued to society by the fraud, social welfare also incorporates the benefits obtained by offenders and is reduced by the costs of catching and sanctioning defendants.

The harm from fraud

Determining the losses attributable to fraud commission is not always straightforward, as fraud is a rather broad concept. In some ways, fraud is analogous to theft. When a healthcare provider bills Medicare for a non-existent procedure, the provider’s gain of money seems like theft at Medicare’s expense. On the other hand, courts have also recognized some forms of regulatory violation as fraud. Under FDA regulations, pharmaceutical companies may not market “off-label” uses for their products. The FDA approves drugs for the treatment of specific conditions, and the companies sell and label drugs for those particular purposes. Physicians, however, may utilize the drugs as they see fit. They may learn or discover other conditions for which the drug is useful, even though the FDA has yet to approve the drug for those other conditions. While the pharmaceutical company may respond to physician inquiries about such off-label uses, they cannot market or advertise these alternative treatments. As an example, Pfizer settled a $2.3 billion off-label marketing case under the False Claims Act in 2009.\(^\text{16}\) While the action may have been a clear violation of FDA guidelines, it is more difficult to measure the social harm from the offense. Perhaps patients who otherwise would have suffered

\(^{13}\) § 3730(d).

\(^{14}\) *Id.*

\(^{15}\) *Id.*

\(^{16}\) See Gardiner Harris, NY Times pg B4, Sept 3, 2009, “Pfizer Pays $2.3 Billion to Settle Marketing Case.”
greatly gained early access to a treatment that was yet to be FDA approved. On the other hand, there may have been patients who were exposed to unnecessary risks and side effects due to insufficient drug testing. We can see similar concerns for other frauds that stem from regulatory violations, such as compliance with the Department of Education’s regulations regarding school loans.

Even in the more direct case of fraud as theft, social welfare analysis might consider the defendant’s situation. Perhaps the defendant is a poor, disadvantaged individual who does not quite qualify for Medicare coverage, yet obtains life saving healthcare through fraudulent billing. Depending on how we value the defendant’s life, it is possible that we might find such fraud socially desirable.

We should also consider the dynamic costs of fraud. Fraud, like theft, is a form of the involuntary transfer of resources. In many ways, the legal system helps facilitate voluntary agreements and transfers in the support of private freedoms. The potential for involuntary transfers can trigger both offensive and defensive investments that seem wasteful. Offenders may expend efforts concealing their fraud, while the government may create additional bureaucracy to slow or complicate attempts at fraud. These costs are above the direct costs of detection and prosecution of fraud.

The efficacy of this system depends greatly on the ability of the judicial system to ascertain the proper costs and, perhaps, benefits, of fraud offenses. If the damages or penalties awarded are out of line with this social calculus, the system can break down. For purposes of the percentage bounty system, fraud penalties must be proportional to the social harm in order to provide the appropriate incentives to relators. Recall that the False Claims Act provides a $5,500 to $11,000 civil penalty per claim in addition to treble damages. A difficult outstanding question is whether courts utilize this sanction scheme to produce total penalties in line with the social costs of the defendant’s actions. I do not address this question here; most of the case values in my data come from settlements. For purposes of analysis, I assume there is no systematic error between social harm and settlement values across the range of FCA settlement values.

Enforcement costs: Private vs public enforcement

In optimizing social welfare, not only does the hypothetical social planner consider the harm from allowing fraud offenses, but the planner must balance the harms against the costs of enforcement. One basic question for enforcement is the possibility of private vs public enforcement. Qui tam litigation is a form of private action in the support of public interests. There is a substantial literature describing such programs generally. Private enforcement of public law has a number of perceived benefits. Private involvement can increase the total resources devoted to fighting a particular problem, and the private parties might be more efficient at doing so. For example, the cost for an employee to monitor an employer might be less than the costs involved if the federal government were to do so. Private involvement might also correct for agency slack; government regulators

---

might be more subject to political pressure or lobbying, but private involvement might shame them into action. Private enforcement could also develop innovations in litigation, settlements, and law.

The potential downsides to private enforcement dovetail with the aforementioned benefits. Private party involvement might generate excessive enforcement—enforcement against parties who should not be liable. Private enforcement might also interfere with the public regulatory system—either interfering with ongoing government efforts, or perhaps triggering even further government slack. Finally, there can be a lack of public accountability for private enforcement.

Private enforcement is of particular interest in the case of fraud due to the enforcement cost problem. As an offense, fraud is unusual in that victims may not be aware of their victimization. When healthcare providers bill Medicare for a non-rendered service, it is not immediately apparent to the Inspector General of Health & Human Services that Medicare has been victimized. Government employees must first investigate claims to determine legitimacy. Fraud detection is thus a costly, involved process that can be burdensome to the government and to legitimate businesses. In comparison, private whistleblowers may have a dramatic cost advantage in identifying fraud in the course of their own work. Of course, private whistleblowers incur a different set of costs for their actions, as their careers may suffer due to future employer perceptions of disloyalty.

More formally, Landes & Posner argue that pure private enforcement, in comparison to pure public enforcement, is theoretically inefficient because of the incentive problem with fines under optimal enforcement.\(^\text{18}\) To minimize costs of detection and enforcement, they recommend a combination of high fine and low detection rate for crime. The high fine, however, will provide an incentive for greater investment by private enforcers, thus increasing the overall social loss.

Polinsky emphasizes the importance of enforcement costs.\(^\text{19}\) Due to a combination of defendant wealth constraints and their ability to cause societal damage, a theoretically high fine and low detection probability may not work. If the costs of raising the detection chance to a level satisfying rational deterrence are too high, public (as opposed to private) enforcement may be the only option. Public enforcement is not constrained by the profit motive of private enforcers. Polinsky also notes that paying private enforcers an amount different from the fine assessed to the defendant may generate socially optimal results.

The percentage bounty: Does it create bad incentives?

The FCA offers a percentage bounty to drive private enforcement. One simple concern is that bounties might induce people to commit fraud, and high bounties might induce higher amounts of fraud. We can imagine a disgruntled employee inducing her employer to commit fraud and then profiting by blowing the whistle. She must be sufficiently


sophisticated to hide her role in the fraud, lest the DOJ cut her off from any reward. Although I do not have any evidence of this fraud-inducing effect, the DOJ’s investigative abilities are important in preventing such abuse.

Another concern is that the bounty system might induce excessive investment in fighting fraud. This is the Landes & Posner (1975) concern with damage multipliers and private enforcement: if private enforcers receive an amount that is greater than the social cost of the offense, they may over-invest in enforcement. Given that the FCA is a treble damages statute, it is possible that the actual social loss due to a case of fraud might be less than triple the base damage amount. Depending on the judicial application of the civil penalty per false claim, actual social losses could be dramatically lower than the penalty. If relators receive a sufficiently high percentage of the defendant payments, there may be incentive for them to over-invest.

Even if the bounty system is not encouraging investment by individual relators, the bounty system still encourages relators to disclose the existence of fraud. The relator may simply discover fraud in the course of her regular work. This whistleblowing behavior can be costly to the relator, and the damage to her career may be irreparable. The fact that a successful relator receives compensation for her efforts may be of some comfort to her, but there is still social loss due to the whistleblowing costs.

The bounty system may also encourage attorneys to invest in searching for relators and cases. If bounties are too high, there may be excessive competition and expenditures among law firms in securing the best relators. Given the prevalence of successful one-shot firms in the data, however, it may be that firms simply search for valuable cases. FCA cases are a possibility, but the bounty system might not play a large role in diverting search resources from other cases.

The Instrumental Role of the Bounty

Establishing a bounty for private enforcement may have a variety of effects upon the occurrence of fraud and the level of enforcement. Although I do not focus on this aspect here, one possibility is that the government’s choice to announce a bounty percentage might directly deter potential offenders: offering a bounty is a method of signaling to potential offenders that the government is serious about enforcement.

In this article, though, I focus upon the instrumental role of the bounty on two parties: the potential whistleblowers and their attorneys. By influencing these parties’ actions, there may similarly be deterrence of potential offenders. In contrast to the aforementioned direct signal, the impact upon potential offenders is mediated through whistleblowers; a high bounty might motivate more whistleblowers to come forward, making it less likely that a potential offender would commit fraud in the first place.

---

For both whistleblowers and their attorneys, one key challenge in interpreting the offer of a percentage bounty is the uncertainty involved. There is, of course, the uncertainty in knowing whether a particular claim will be successful. More importantly, though, it is unlikely they have good knowledge as to the reward she will obtain the future. The government is not offering a fixed dollar amount for coming forward. Rather, the eventual payment is tied to the value of the claim, and the whistleblower may have some vague notion as to the approximate value of her claim. It will be up to the government, the attorneys, the defendant, and the judicial system to sort out the claim’s value, after which the whistleblower will then obtain her bounty percentage. To the extent we rely on the individual whistleblower to make an informed decision as to her own best interest, she would have to calculate the expected benefits by multiplying the bounty percentage by the value of the claim and the probability of success.

Because of this high level of uncertainty and the fact that most whistleblowers are not repeat players, it is difficult to believe that whistleblowers make an independent, rational decision as to the expected payoffs given a bounty percentage. More likely, potential whistleblowers are not even aware of the False Claims Act and the attendant bounty percentages. Their primary source of information is probably their attorney, once they decide to take an initial step in potential litigation. To the extent an attorney is a faithful agent or representative of the whistleblower’s interests, the whistleblower might make a more rational decision by following the attorney’s guidance. Of course, the attorney may not have an accurate estimate of the value of a claim, but the attorney’s nature as a repeat player might allow for better estimates on average.

The bounty percentage could have a parallel direct impact on attorneys. Even if they are not explicitly acting as the whistleblower’s faithful agent, the fact that the attorney works on a contingency percentage helps align the attorney’s interest with larger bounties. Based on this expectation of receiving more money at the end of litigation, higher percentages could motivate attorneys both to move forward with particular cases of fraud and to invest in discovering new information.

This instrumental view of the bounty percentage has some interesting complications when viewed from the social planner’s perspective. One basic assumption a social planner might make would be that relators and their attorneys are properly informed about the potential costs and benefits of participating in the qui tam process. This is not a necessary assumption for an instrumental view. Nonetheless, this assumption allows the social planner to focus less upon the impact of private enforcement on the relators and their attorneys. Since fully informed relators and attorneys would not move forward if they did not expect to be in a better situation, the social planner can assume that relators and attorneys are better off having participated. On the other hand, if relators are not fully informed about the costs of becoming a whistleblower, this might be a concern for a social planner. Just because a relator decided to participate given a specific bounty percentage ex ante does not mean that she would do so again given the chance. Relators may feel regret given how they are treated; they may be underinformed prior to making their decision about whistleblowing. As a result, they may not actually be better off after choosing to become a relator, even if the eventual payout is as promised. Worse yet, even
if relators are properly informed ex ante, their preferences might change as a result of exposure to the system, making a social planner’s job more difficult.

If we set aside these concerns about ex-post relator well-being, an instrumental social planner still needs to address repeat-player concerns. To the extent that relators are repeat players, having outcomes not match initial relator expectations can severely decrease relator participation. On the other hand, if relators are typically one-shot players, the instrumental social planner might not be concerned. It is also possible that the high costs to becoming a relator is primarily suffered the first time around: for example, a relator’s career prospects may be severely harmed after she becomes a first-time whistleblower. Her decision to whistleblow in the future, however, might not further harm her career significantly. Therefore, she might actually be more willing to whistleblow in the future.

Indirect impact upon attorneys and whistleblowers

The impact of establishing a percentage bounty for private enforcement might be less direct than the rational expectation approach described above. Rather than attorneys or whistleblowers calculating their expected rewards, the percentage bounty might impact these parties via more diffuse methods.

One example might be their reliance upon actual announcements of awards for resolved cases. Under this mechanism, rather than relying upon the legislative bounty percentage, whistleblowers and their attorneys rely upon public announcements of actual resolved cases. Thus, a change in the bounty percentage is only effective to the extent it changes actual case rewards. This model has a number of implications. First, any legislative change in the bounty percentage might not have any immediate impact, as whistleblowers wait to see how real cases are resolved in the legal system. Second, it is possible that courts or other actors might circumvent a legislative change in the bounty by using discretion. If, for example, the bounty percentage were to be reduced, courts might find certain cases unfair to the whistleblower and adjust total recoveries upwards such that the whistleblower would receive the same total dollar amount despite the legislative bounty percentage change.

As another alternative, potential whistleblowers might be more motivated by their knowledge of other whistleblowers. To the extent that the bounties motivate attorneys to invest more in bringing cases forward, whistleblowers may be more willing to come forward simply because they are more likely to know or hear about other whistleblowers. On a larger scale, the knowledge of the bounty and of other whistleblowers might slowly shape social mores or beliefs about the acceptability, and perhaps even obligation, of whistleblowing. These mechanisms would similar take substantial time; we would not expect a legislative change in the bounty to incur immediate impacts.

The Rational Actor model is relatively common
Numerous papers either explicitly or implicitly take this rational relator approach, in which relators are positively motivated by greater bounties. Bucy presents a complex game theoretic model incorporating the relator, the defendant, and the regulatory as three players.\textsuperscript{21} Although not explicitly in her model, she does mention the fact that the potential relator's counsel is a repeat player, thus justifying the design of her analysis in the form of repeated/iterated games.\textsuperscript{22}

Depoorter & De Mot also present a game theoretic model of the FCA, utilizing the same main three players as Bucy.\textsuperscript{23} Their model emphasizes the differences in probability of success between the relator acting alone versus the government intervening. With a government actor that values the recovered dollars, it predicts that the government will intervene in high dollar cases along with low dollar cases that the relator would not otherwise pursue. They hypothesize that this setup may trigger underprovision of qui tam cases if the relators recognize the government's potential for free riding. Finally, they mention the perverse incentive for relators to delay in obtaining maximum damages.

Heyes & Kapur, although not specific to the FCA, present an economic model of the whistleblower.\textsuperscript{24} They highlight different decision metrics by which whistleblowers decide to take action along with the impact of "noisy" or imperfect information.

Since relators receive a percentage of the damages, various commentators argue that the relators have an incentive to allow the fraud to increase, thus improving the relator's reward. In a related strain of argument, others are concerned that relators target relatively trivial contractual violations, and then pursue claims of fraud for the value of the entire contract (or series of contracts), thus unjustifiably penalizing the defendant.\textsuperscript{25} Many of the statutory changes of the FCA, along with commentator proposals, thus have centered on restructuring the maximum bounty percentage. The FCA originally guaranteed a full 50% of recovery to the relator.\textsuperscript{26}

Brollier suggests a graduated bounty percentage that will encourage prompt revelation.\textsuperscript{27} His proposal grants up to 30% to relators who file within two years of defendant's first fraud, decreasing down to 10% for relators who file after four or more years have elapsed. In contrast to Brollier's suggestion, Kovacic proposes that the relator's damage award be constrained by the time at which the relator knew or had constructive knowledge of the misconduct.\textsuperscript{28} Trunk recommends improved incentives for voluntary

\textsuperscript{22} Bucy at 627.
\textsuperscript{23} Ben Depoorter & Jef De Mot, Whistle Blowing, 14 Supreme Court Economic Review 135 (2005).
\textsuperscript{24} Anthony Heyes and Sandeep Kapur, An Economic Model of Whistle-Blower Policy, JLEO 2008.
\textsuperscript{26} James B. Helmer, Jr., How Great Is Thy Bounty: Relator's Share Calculations Pursuant to the False Claims Act, 68 U. Cin. L. Rev. 737, 739 (2000).
\textsuperscript{27} Brollier (2006) (at 716-18)
\textsuperscript{28} Kovacic (1996), 29 Loyola L. Rev 1799, William E. Kovacic, Whistleblower Bounty Lawsuits as Monitoring Devices in Government Contracting
disclosure of compliance/contractual violations, including assurances of continued
government business and guarantees of less than treble damages.\textsuperscript{29}

Ferziger & Currell make broad recommendations of low (single digit percentage)
bounties, maximizing informant anonymity, and increasing the predictability of bounty
payment.\textsuperscript{30} Their ideal bounty percentage "should be equal to the agency's average
information cost per dollar of enforcement revenue, not including the bounty program's
marginal operating cost."\textsuperscript{31}

These commentators recognize specific concerns within the private enforcement scheme,
and speculate on potential impacts. In contrast, I address two simpler questions with
systematic empirical evidence. First, what is the impact of the current bounty percentage
on government interests? Second, what would be the result of a change in the bounty
percentage?

A Simple Model of Private Enforcement

To answer these questions, I begin with a simple model of private enforcement. The three
parties are the DOJ, the potential defendant, and the potential relator. Working from a
rational crime perspective, the potential defendant is a profit maximizer. Since fraud is an
economic offense, the profit maximizing assumption is more reasonable.\textsuperscript{32} He will
commit fraud if he expects it to be profitable. If the probability of being caught is too
high, the DOJ has successfully deterred him from committing fraud. I am not explicitly
considering non-economic criminal sanctions, and I start off analysis as a one-shot game
for potential defendants and relators.

The potential relator is an individual who has knowledge of fraud otherwise unknown to
the DOJ. She recognizes that becoming a relator is a costly activity, as whistleblowing
can be a lengthy process with serious career consequences. She may have various
motivations in becoming a relator, including personal satisfaction in seeing wrongdoing
exposed. The important assumption in this simple model is that she responds positively to
financial reward. A larger bounty percentage increases the probability that she will step
forward and become a relator. Her decision is a binary decision: either she blows the
whistle, or she does not. She is honest in that she neither causes fraud nor falsely blows
the whistle when there is no fraud; she only triggers false negatives by not blowing the
whistle when there actually is fraud. The relator is typically represented by counsel, but
initially I set aside potentially conflicting incentives by assuming that counsel is a faithful
agent to the relator principal.

\textsuperscript{29} Trunk, at 177
\textsuperscript{30} Ferziger & Currell, Snitching for Dollars: the Economics and Public Policy of Federal Civil
\textsuperscript{31} Ferziger & Currell, 1999 U. Ill. L. Rev. at 1187.
\textsuperscript{32} I similarly assume risk neutrality at this point. Although I treat fraud as a binary decision, there is
obviously a wide variety of methods of fraud that I abstract to an aggregate dollar value.
For now, I also abstract away from the remaining qui tam procedure. Once the relator decides to file, the DOJ handles the rest. Thus, this simple model is analogous to other reward schemes in which enforcement agencies offer cash for information leading to arrest or prosecution. While there are many methods the DOJ might use to encourage potential relators to take action, I here look at the DOJ's choice of bounty percentage.

In designing its fraud enforcement scheme, the DOJ has at least two priorities: compensation and deterrence. The defrauded government agency desires compensation for its loss. This principle of compensation is more precisely described as net government compensation, since paying the relator reduces the amount of government compensation. Furthermore, it may be better if the agency had never suffered loss in the first place. A sufficiently strong enforcement scheme might deter potential offenders. The public litigation bounty is one method by which the DOJ can encourage relator participation in the enforcement scheme.

I begin with a DOJ that focuses only on government compensation. With this objective, it wishes to maximize recovery dollars going to the government from defendants. The bounty paid to relators is a cost that reduces net government recoveries. As such, this DOJ will set a bounty percentage high enough to encourage relator participation, but not so high that relators receive all of the recoveries. More precisely, this DOJ will select a bounty percentage such that the marginal gain due to the last case brought forward is equal to the marginal cost of paying the bounty. If the DOJ were to raise the bounty any higher than this level, it would encourage another relator to come forward with a new case, but the gain in prosecuting that case would be outweighed by the additional bounties paid to all of the relators. If the DOJ were to set the bounty lower than this level, it would be “leaving money on the table.” In other words, by raising the bounty percentage, relators would bring new cases forward. The recovered dollars from those cases would outweigh the additional bounty percentage paid for all cases.

A DOJ that focuses solely on deterrence would behave differently. Since a higher probability of detection makes potential defendants less likely to commit fraud, it could attempt to maximize the probability of detection. One simple solution would be to maximize the bounty percentage. If the DOJ does not spend its own funds, the maximum bounty percentage would be 100%. The success of this system approach depends on the susceptibility of potential defendants to detection and thus deterrence, though, along with the responsiveness of potential relators. It is possible that despite high bounty percentages, potential relators might be generally unwilling to come forward. As a result, although the 100% bounty might maximize their willingness, the still low probability of their action might not generate much deterrence. If, for example, most frauds result in a government loss of about $10,000, we might few relators to come forward for a 100% bounty because they place a much higher value on their careers than a potential $10,000 one-time payout. Following a similar line of logic, if many potential defendants believe their probability of detection is low even at a 100% bounty, a 90% bounty might effectively produce the same level of deterrence.
Now, I consider a DOJ that incorporates both deterrence and compensation objectives. Such a DOJ would likely choose a bounty percentage in between the aforementioned levels. More precisely, the DOJ would select a bounty percentage such that the marginal compensation benefit is equal to the marginal deterrence benefit. The compensation-only DOJ previously selected a percentage at which a higher percentage would have resulted in a net compensation loss for the DOJ. By factoring in the additional deterrence gain from the higher percentage, though, the unified objective DOJ can find the higher percentage acceptable.

It is possible, though, that potential defendants simply are not deterred at the relevant bounty percentages. If this is the case, the unified objective DOJ would behave like the compensation only DOJ. There is no additional deterrence gain in raising the bounty percentage, so it only considers the compensatory value in setting the bounty.

First Implication of the model: Deterrence disagreements

In comparing the two goals of deterrence and compensation, it is important to note that compensation is easier to observe. We can measure the total dollars recovered in a rather straightforward manner, but properly measuring the amount of potential fraud deterred can be challenging. How can we accurately determine how much fraud would otherwise have occurred? To an outside observer disregarding this comparative measurement problem, a DOJ establishing a relatively high bounty percentage may appear to be “irrationally” paying too much to relators. The argument would be that the relators who came forward with information under the high bounty percentage would still have done so under a slightly lower bounty percentage. Therefore, the DOJ is being wasteful by paying too much for information and thereby failing to maximize net compensation.

Following the broader line of reasoning in this section, however, the discrepancy may be explained by differences in estimating deterrence. The DOJ might believe that the higher bounty percentage generates substantial deterrence, while the outside observer might disagree or not even value deterrence.

Second Implication: Administrative ease of fixed bounty percentage

Another important implication is that a fixed bounty percentage has the advantage of administrative ease. In an idealized sense, a DOJ focused exclusively on compensation might want to engage in price discrimination among relators. Such a DOJ might try to determine the minimum it would have to pay any particular potential relator to induce her to provide information. This theoretical price discrimination scheme would be complex and administratively difficult, as each relator would be reluctant to disclose her price flexibility. Offering a fixed bounty percentage makes it easier for relators to set expectations regarding their personal compensation for whistleblowing. Making the decision to become a whistleblower has serious career and social consequences; the prospect of also having to negotiate with the government regarding bounty percentages may not be attractive to a potential relator.
Besides administrative feasibility, however, note that a fixed percentage may also help drive deterrence against larger cases of fraud. If the magnitude of fraud offense varies, a fixed bounty percentage may inadvertently grant “too much” to a potential large-fraud relator from a compensation perspective. Nonetheless, the large bounty may help increase the probability of relator participation and thus drive deterrence of such large frauds.

Third implication: Compensation and deterrence are only tradeoffs at the high end.

The third implication is that compensation and deterrence are not always in competition as goals. If the current bounty percentage is particularly low, increasing the percentage may result in both more compensation and more deterrence. On figure 1, this would be zone A. If the government can determine that its current bounty percentage is in zone A, increasing the bounty is a win-win. Only after the current bounty percentage exceeds the compensation maximum, zone B, do we actually confront a tradeoff between compensation and deterrence. Within zone B, the government must then determine how much it values deterrence in comparison to compensation for fraud. If deterrence is more important than compensation, it may select a bounty percentage on the right side of zone B. If deterrence is difficult to measure or less important than compensation, the government may end up choosing a spot closer to the left side of zone B.

Fourth implication: A limited view of deterrence

Also note that this model has a very limited approach towards deterrence. There are many possible ways potential defendants might be deterred from committing fraud that do not involve prosecution. The government might learn of a potential area for fraud and publicize its awareness; even if no one has yet to be prosecuted for that type of offense, potential offenders may be deterred. The government might also deter offenders by not granting future contracts or applying unfavorable terms. Furthermore, this model of deterrence assumes that potential offenders actually learn about the probability of being caught. Their perception of this probability might function less by volume of cases caught, but more by the number of high profile cases publicized. Seeing a large dollar prosecution of a large corporate entity might have greater relevance for both media outlets and other like-minded offenders, whereas smaller defendants might not attract similar publicity.

Figure 1
IV. The empirical challenges

From a policy perspective, the goal is to determine the proper bounty percentage. There are a number of empirical challenges in that determination: I will only address two here. First, we need a better measure of performance: rather than just total recoveries, we should consider what percentage of fraud is being caught. Second, how will potential whistleblowers respond to a particular bounty percentage?

By taking the aforementioned instrumental view, the empirical challenge is in predicting how potential whistleblowers will respond to a particular bounty percentage. One straightforward method is to survey people. Feldman & Lobel make important contributions by conducting laboratory survey experiments, and they consider a variety of factors besides money in the individual decision-making process. Asking individuals whether they would blow the whistle given a hypothetical scenario is one method of predicting whether they would actually do so in reality. Given the severe social and professional repercussions of whistleblowing, though, we may be concerned about the external validity of hypothetical survey results. Various studies suggest a weak

---

33 See Mesmer-Magnus and Viswesvaran, Whistleblowing in Organizations: An Examination of Correlates of Whistleblowing Intentions, Actions, and Retaliation, 62 J. of Business Ethics 277 (2005) for a review of such literature.

relationship between intention and actual action.\(^{35}\) It may be difficult for individuals to respond truthfully to a survey as to their willingness to whistleblow if they have not personally encountered the costs of doing so.

Another alternative is to survey actual whistleblowers. These are individuals who actually have faced the real repercussions of whistleblowing; their opinions should perhaps be due greater deference. There are concerns, though, that actual whistleblowers may censor their responses due to lack of anonymity and self-selection.\(^{36}\) It may be that actual whistleblowers are systematically different from non-whistleblowers. As a result, we may be unable to extrapolate from their answers in surveys because non-whistleblowers would not value the same things. For example, if actual whistleblowers uniformly answered surveys indicating that monetary reward was not a factor in their decision, we might be suspicious of inferring that monetary reward would not induce whistleblowing in other people.

Rather than conducting surveys, another alternative is simply to observe whistleblower behavior. This reduces the risk that individuals might say one thing and do something else, as this form of observation relies on whistleblower actions. One of the downsides to relying on pure observation is that we have limited control over variation and variables of interest. Unlike surveys based on hypothetical situations that we can create ad nauseam, for example, the observed whistleblowers face the circumstances dealt by real life.

Relatedly, observational data results in a serious causal inference challenge. This article’s focus is on setting the proper bounty percentage. We actually observe different bounty percentages being paid out; after resolution of FCA cases, courts have some discretion in granting a bounty percentage between 15% and 25%. There are a multitude of factors that can go into this determination, many of which may not be clear to the whistleblower at her time of whistleblowing decision. If we take our instrumental view of the bounty percentage, though, it is hard to believe that we can use this variation in bounty percentage to predict how potential relators would act. Being paid 17% after resolution of a case is different than expecting to be paid 17% before bringing a particular case forward.

One solution to this causal inference challenge is to find a source of variation in the bounty that is exogenous, or, in other words, not properly attributable to other relevant factors in the decision to whistleblow. This form of variation is analogous to an experiment, in which nothing else changes besides the variable of concern. This type of observational data would then allow us to infer that changes in the observations are due to the experimental variation. Restated, we would be able to infer that changes in the rate of whistleblowing are due to this variation in the bounty percentage.


\(^{36}\) See arguments presented in Mesmer-Magnus at 278-79.
I should note a broader empirical concern with this instrumental emphasis on relator action, which is the normative application. To some extent, a social planner might use this instrumental view to set a bounty percentage: given this knowledge about relator performance, a social planner could set a percentage to induce a level of relator participation. If the social planner wants to induce greater participation, it can increase the bounty, and vice versa. On the other hand, we might be concerned about the social costs of relator participation, particularly to the relator herself. Just because a relator decided to participate given a specific bounty percentage ex ante does not mean that she would do so again given the chance. In other words, while some of these methods may generate revealed preferences at the time of bounty offer, her preferences may change as a result of her exposure to the qui tam system. Relators may feel regret given how they are treated; they may be underinformed prior to making their decision about whistleblowing. As a result, they may not actually be better off after choosing to become a relator, even if the eventual payout is as promised.

V. Performance of the current bounty.

Empirical evaluation of a fraud enforcement system is difficult due to the challenge of offense detection. Unlike more physical crimes such as homicide, the victim of fraud often may not realize she has been victimized until the offender is caught. Without background victimization information, we cannot make claims that a program has prosecuted 80% of fraud offenses. The volume of observed offense prosecution could reflect a complex interaction between the unobserved offense prevalence and enforcement efforts.

In healthcare, HHS has attempted to estimate the total annual amount of “improper payments” under Medicare and Medicaid, and some have touted the 2010 estimate of $70 billion as a background fraud estimate. HHS generated the improper payments estimate through random sampling of medical claims. If the provider was unable to support a claim with documentation, for example, HHS would flag the claim as an improper payment. Due to the tremendous volume of healthcare claims, this technique is useful in estimating the error rate in payments, but it both over and under-identifies fraud. For over-identification, we typically view fraud as a purposeful action with an attendant mens rea. Sloppy or misfiled documentation may simply be a mistake or negligence not rising to the level of fraud.

The under-identification problem is more serious, though, from a performance measurement perspective. First, documentation is notoriously easy to manufacture. For example, consider practices such as “up-coding” or “up-charging,” in which medical providers intentionally diagnose patients with worse conditions than merited for billing purposes. If the practice is common and subtle, it should be easy to maintain documentation in compliance with the exaggerated diagnosis. Ex post verification of whether the patient actually had acute chest spasms or just a mild cough would be extremely difficult. Second, the HHS estimate does not include some of the major fraud
recoveries under the False Claims Act, such as off-label advertising fraud\textsuperscript{37} or average wholesale price litigation\textsuperscript{38}. As such creative methods of fraud come to light, we lack systematic evidence of the underlying frequency of such frauds.

An estimate of the FCA detection rate

The approach I take in estimating enforcement performance focuses on the relationship between fraud offenses of varying magnitude. I claim that the underlying minimum volume of smaller offenses can be approximated by the volume of observed larger offenses. For example, if I observe 30 cases of $10 million fraud in one year, I expect that there are at least 30 cases of $1 million fraud occurring in that time period.

This expectation can be justified via two parallel methods. First, to the extent that offenses are an aggregation of repeated fraudulent behavior, it is possible that the $10 million fraud might have been detected earlier. An unscrupulous healthcare provider might steadily increase its monthly volume of fraudulent Medicare claim. At an earlier point, the accrued fraud might only have reached $1 million. It was only the failure to detect and stop the fraud earlier that resulted in the $10 million value of the offense. Thus, the observation of the $10 million offense reflects an offense that was unobserved when it was $1 million in value. This assumes, of course, that all of the aggregated offenses fall within the appropriate statute of limitations.

The second justification is based upon a prediction of defendant behavior. Most enforcement schemes apply greater penalties to large offenses over smaller offenses. The FCA is no different, as criminal sanctions generally apply only to the worst offenders, and civil penalties scale with the magnitude of the offense under the treble damages rule. Under this type of enforcement scheme, I predict greater deterrence of worse offenses over smaller offenses. Unless greater resources are targeted explicitly at the smaller offenses, the offense volume for those smaller offenses should be at least at the level of the worse offenses, if not greater.

Utilizing this approach to estimate the background level of offenses, we can look at the distribution of offenses by offense value.

Figure 2: Histogram of prosecuted offenses by log offense value

\textsuperscript{37} Pfizer settled a $2.3 billion off-label marketing case under the False Claims Act in 2009. See Gardiner Harris, NY Times pg B4, Sept 3, 2009, “Pfizer Pays $2.3 Billion to Settle Marketing Case.”

Figure 2 shows the volume of successfully prosecuted offenses, with a peak between $e^{13}$ and $e^{15}$ ($440,000$ and $3.2$ million). I will call this range of offenses the middle value offenses. On either side of this peak, the volume of offenses decreases.

Utilizing this distribution of offenses, I infer that the FCA is less successful at prosecuting offenses valued under $440,000$ relative to offenses between $440,000$ and $3.2$ million. If, hypothetically, the DOJ were catching and prosecuting 50% of middle value offenses, I would claim that the DOJ caught less than 50% of the sub-$440,000$ offenses. An alternative interpretation, of course, is that there are simply fewer offenses occurring that are valued under $440,000$. For the above reasons, I believe this is unlikely. Nonetheless, it is possible that potential offenders are less willing to commit smaller amounts of fraud. Perhaps, for example, their fear of criminal or other sanctions is so great that only large amounts of money are sufficient to compensate for the risk.

This information does not help much in understanding the performance of the FCA in the $440,000$ to $3.2$ million zone. To the right of the $3.2$ million, the rate of prosecuted offenses declines. This might suggest that deterrence is at play, in that large bounties or criminal sanctions reduce the probability of committing high dollar frauds. Since the high dollar fraud commission rate is lower, there are simply fewer cases to be caught. On the other hand, it could also be evidence that defendants with the resources to commit such large frauds also fight more vigorously against detection and sanction. Alternatively, judges and juries might be unwilling to award such large sanctions for fraud despite the prevalence of such actions.
For purposes of this evaluation, though, it is sufficient to say that there appears to be opportunity for improvement in prosecuting cases valued under $440,000. The next question is whether relators could play a role in the improvement. Given a higher bounty percentage, would relators be more willing to bring these sub-$440,000 cases forward?

VI. Measuring responsiveness of relators to the bounty percentage

Predicting how relators react to various bounty percentages is difficult. The ideal situation would be to have a randomized experiment as to the effective finder's bounty percentage. We could then make strong recommendations as to the appropriate bounty percentages.

Unfortunately, not only do we lack a randomized experiment, we do not have much variation in the finder's bounty percentage. Since the FCA's 1986 amendments, the base finder's bounty percentage has remained constant at 15%. The lack of variation on the bounty percentage makes it difficult to observe how relators would respond to different bounty percentages. The statute provided different bounties prior to 1986, but the 1986 amendments had other significant changes besides the bounty percentages. It would be challenging to determine which portion of increased relator participation would be properly attributable to the bounty percentage change as opposed to the other statutory changes. Further clouding the issue would be the broader differences in government contracting, cultural norms, and legal environment before and after 1986.

My strategy is to look at a source of exogenous variation in the effective bounty percentage. The American Jobs Creation Act of 2004 (AJCA) changed the rules for the tax treatment of plaintiff awards paid on or after October 23, 2004. Prior to that statute, the Internal Revenue Service held that the plaintiff's proceeds were fully taxable as income, including the attorneys' fees paid under a contingency arrangement. Although the attorneys' fees could be deducted as a miscellaneous itemized expense, this deduction was not available to taxpayers subject to Alternative Minimum Tax (AMT). As a result, some plaintiffs who paid large contingency fees to their attorneys would take home very little of the total award because the IRS taxed the plaintiff as if she had not incurred any attorneys’ fees. After the AJCA, the plaintiff's net proceeds after attorneys' fees were treated as income, thus avoiding the AMT and deduction problem. Under the AJCA tax treatment, plaintiffs could now be assured of paying income taxes only on their take-home award.

I propose that the AJCA is a source of exogenous variation as the motivations for passing the law do not seem related to the underlying causes of qui tam cases. As a contrasting example, consider a hypothetical judge whom we observe to grant higher bounty percentages in comparison to other judges. We would be suspicious of drawing inferences from this difference in this judge's bounty awards. Her award of greater percentages might be the result of particularly reprehensible defendant conduct in her jurisdiction. She could also be an unusually demanding judge, such that only the most skilled relators would appear before her, yet she would compensate for the high demands by offering higher bounties, too. Utilizing a source of exogenous variation reduces the impact of other causal arguments in predicting the impact of the bounty change.

The impact of this tax treatment change is an effective increase in the bounty percentage, but it is not a uniform increase for all relators. Some relators would not experience any increase in the bounty: for example, a relator who was not subject to AMT and would have itemized deductions regardless of litigation. There are two groups of relators who do receive an effective increase. The first group consists of relators subject to AMT. Given that the median relator share is approximately $144,000, many relators likely fall into this category. For large rewards, the change is uniform, as the reward itself is sufficient to lift the taxpaying relator's income into the AMT range. For smaller rewards, however, the impact depends upon the non-qui tam income of the relator. If the relator's total income was sufficient to subject her to AMT, then the tax treatment change provides a higher effective bounty.

The second group of relators who receive an effective increase are those who did not benefit from the miscellaneous itemized deduction of attorneys' fees. These are relators who did not have enough deductions to justify itemizing deductions and would have taken the standard deduction on their income tax. It is possible that the miscellaneous itemized deduction of attorneys' fees was large enough to justify itemization, but on the margin, the benefit might not be that much greater than the standard deduction they would have received. Of course, if attorneys' fees are lower than the standard deduction, the relator would receive no marginal tax reduction under the pre-AJCA regime. The AJCA treatment allows these relators to benefit from both the standardized deduction and the removal of the attorneys' fees from income. Given that the standard deduction is on the order of $5,000 to $10,000, it is unclear how many relators would fall into this category. A qui tam case must be rather small if the standard deduction significantly weighs against the importance of the attorneys' fees. Without more detailed information regarding relator income tax filings, it is difficult to estimate the size and impact of this tax treatment.

My rough prediction, then, is that the effective shift in tax treatment should disproportionately favor larger cases. These larger cases will result in larger bounties that subject relators to AMT. Unless relators are uniformly high income individuals subject to AMT, or if relators filing low-dollar cases are distinctly higher income, it is more likely that the relator filing a high-dollar case would benefit under the AJCA tax regime.
To estimate the magnitude of the tax treatment improvement, consider that the AMT is roughly 28% on income over $175,000. Next, I approximate the contingency fee for the attorney at 40%. A relator subject to AMT receiving a bounty before October 23, 2004, would not only have to pay her attorney 40% of her share, but she would also pay 28% of that 40% to the IRS, which translates to an additional 11.2% of her total share. I ignore the tax she owes on the original 60% of her share, since that does not change under the AJCA. Thus, she takes home 48.8% of her share of the qui tam recovery. After October, 2004, she can now take home the 60% of her share, as she is no longer responsible for the 11.2% under AMT. This is approximately a 23% increase in the amount she takes home after October 23, 2004.

First stage: relator recognition of improved tax treatment

The first stage in measuring the elasticity of relators to greater financial compensation is potential relator recognition of the effectively greater reward under the AJCA tax regime. Before a potential relator decides to file, she must weigh various factors of taking action, including perhaps the costs and benefits of proceeding with a qui tam claim. Becoming a whistleblower can have dire career and social consequences, and others might view a whistleblower receiving financial reward to be particularly distasteful. To utilize the AJCA as a source of exogenous variation, it must have some cognizable impact on the relator’s decision. This challenge is compound, though: it is unclear how well whistleblowers can accurately measure any of the aforementioned costs and benefits. Even if a whistleblower can approximate the potential financial rewards of a qui tam action, she must then consider the tax consequences of the qui tam bounty, since receipt of litigation proceeds is probably an unusual tax circumstance for most people.

One potential vector of either information or incentive is the relator's attorney. I acknowledge that relying upon an attorney removes the effectiveness of the AJCA variation upon potential whistleblowers who never even see an attorney in the first place. On the other hand, there are many potential whistleblowers who do have initial meetings with attorneys but do not move forward with litigation. As long as these potential whistleblowers take the first step in contacting an attorney, the AJCA variation could impact the relator’s decision to move forward. A responsible attorney might discuss the tax consequences of potential rewards with the relator in helping her make a decision about proceeding with a qui tam filing. Another possibility is that the attorney works harder or is more likely to encourage a relator to move a claim forward under the AJCA tax regime because he predicts that the relator will be more satisfied with the eventual financial reward. We might imagine such an attorney to be a faithful agent to the relator principal in understanding the relator’s incentives.

Note that the relator’s attorney does not officially benefit from the favored tax treatment under the AJCA, except to the extent that it encourages greater relator participation on either the extensive or intensive margins. Nonetheless, I do not have information as to the specific contingency fee arrangements made between relators and their attorneys. It is possible that attorneys could have appropriated the relators’ tax benefits by increasing their contingency fees by a corresponding amount. Stated another way, law firms may
have a policy of charging a fixed contingency percentage for handling a case, but not
every law firm charges the same percentage fee. A law firm charging a higher
contingency percentage might become more aggressive in recruiting relators in response
to the AJCA treatment, leading to generally higher contingency fees.

Again, this first stage is extremely difficult to evaluate; it is unclear if potential relators
and their attorneys have much precision or accuracy regarding the value of their
information. The improved tax treatment could generate an upward nudge in their
preliminary value estimate.

The second stage: Relator responsiveness to an increased reward

Assuming that knowledge of the improved AJCA tax treatment does find its way to the
potential relator and her attorney, I now consider the impact of the effectively increased
reward. Under the finder's bounty, the financial incentive is for relators to bring
information to the DOJ. The ostensible prediction is that a relator is more likely to come
forward if the bounty percentage is greater.

The AJCA is quite clear that cases paying out after October 23, 2004, should be subject
the more favorable tax regime. Assuming that the relator or her attorneys are aware of
this improvement in tax treatment, we might expect the change in relator filings to occur
immediately after the October date. Complicating matters, however, is whether or not
relators might have had earlier knowledge of the AJCA. It is possible that they might
have been aware of the Act before October. Given that the expected time under seal while
waiting for the DOJ's election is over a year, such relators could have increased their rates
of filing before the October 2004 date. This anticipatory filing behavior would weaken
my estimate of relator elasticity using the October date as a discontinuity. Alternatively,
relators could react to the shift in tax treatment by holding off filing qui tam actions until
immediately after the October 2004 date. This type of relator would have otherwise filed
earlier, but decided she would rather obtain the guarantee of an effectively higher bounty
percentage. This alternative anticipatory action would bias the discontinuity estimate
upwards, since the “additional” filing generated by the relator after the October 23, 2004
date would not be a marginal case in the counterfactual pre-October 2004 situation.

The extended time pending DOJ election causes a further challenge in evaluating relator
elasticity. The FOIA data end in July, 2009. Cases spend an average of 600 days under
seal, with a median time of 437 days and a standard deviation of 517 days. Given this
lengthy time under seal and my lack of data regarding sealed cases, we likely observe
only a small fraction of cases filed in 2007 or 2008. Since the key filing date is October
23, 2004, I do not have many years of final outcome data after that date. Without
substantial post- October 23, 2004 data, a regression discontinuity design does not seem
feasible until the DOJ provides further information.
Qui tam cases per year (year in which the attorney general was served)

Given the sparse data I have presently, I utilize a difference approach. I look at cases filed in the years immediately before and after October 23, 2004. I will refer to these as pre-AJCA and post-AJCA years.

<table>
<thead>
<tr>
<th>Years</th>
<th>n (cases with values)</th>
<th>Median Log Case Value</th>
<th>Mean log case value</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2 (adjusted 2003)</td>
<td>59</td>
<td>14.56</td>
<td>14.71</td>
</tr>
<tr>
<td>-1 (adjusted 2004)</td>
<td>46</td>
<td>13.8</td>
<td>13.63</td>
</tr>
<tr>
<td>1 (adjusted 2005)</td>
<td>34</td>
<td>13.73</td>
<td>13.78</td>
</tr>
<tr>
<td>2 (adjusted 2006)</td>
<td>20</td>
<td>13.27</td>
<td>13.54</td>
</tr>
</tbody>
</table>
Median log case value vs year (year in which AG was served)

Twoway histogram: density of case values. Red is two years pre-AJCA, Navy is two years post-AJCA
There is a decrease in the median case value after October 23, 2004; above I show a graph overlaying the distribution of case values before and after the AJCA date. This suggests that the increased incentive is disproportionately affecting the smaller value cases. Alternatively, the reduction in median case value could be driven by a lower relative frequency of high dollar cases, but it would be difficult to attribute this to the preferential tax treatment.

From a rough visual analysis, the greatest region of relative increase is between $e^{10}$ and $e^{12}$, which roughly corresponds to $22,000 and $160,000. Cases around $e^{13}$, around $440,000$, also seem to rise in proportion, but to a smaller extent.

From this evidence, I infer that for large dollar cases, relators were already moving forward in the pre-AJCA regime. The additional effective bounty they receive in the post-AJCA regime does not impact them in the same way as relators considering small dollar cases. The potential relators with knowledge of small dollar cases in the pre-AJCA regime were relatively less likely to participate in the qui tam system. After the increased bounty, some of those reluctant potential relators seem to have moved forward with their qui tam claims.

Verification

The challenge with this approach is that it is entirely possible that something else may have fundamentally changed around the October, 2004 timeframe that would have resulted in this effect. My assumption in measuring the responsiveness of relators to the AJCA change is that other factors related to qui tam actions would have remained smooth for purposes of regression. The difference approach I use is even more rough than a regression discontinuity design. Nonetheless, I can at least verify that other case attributes are relatively consistent before and after October 23, 2004.

<table>
<thead>
<tr>
<th>Settlement/Judgments</th>
<th>All cases</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre-AJCA (2yrs)</td>
</tr>
<tr>
<td>n</td>
<td>105</td>
</tr>
<tr>
<td>Primary Agency = HHS</td>
<td>64 (61%)</td>
</tr>
<tr>
<td>Intervention rate</td>
<td>87%</td>
</tr>
<tr>
<td>Median Time to Decide (days)</td>
<td>907</td>
</tr>
<tr>
<td>Primary Agency = DOD</td>
<td>13 (12%)</td>
</tr>
<tr>
<td>Dismissal rate</td>
<td>n/a</td>
</tr>
</tbody>
</table>

The above table considers two years of cases immediately before and after the AJCA date. The first and second columns reflect characteristics of the cases in which there was an imposition against the defendant; these are the only cases I consider above, as I have no other method of distinguishing case values. The third and fourth columns include all unsealed cases filed during the two years before and after the AJCA date. Many of these comparisons suffer simply from the small volume of cases. Nonetheless, there do not appear to be dramatic differences before and after the AJCA date. The reduction in median time to decide regarding election supports the censoring difficulty. That is to say, likely many cases are not showing up in the years after the AJCA because they remain under seal.

What is the counterfactual for the sub-$440,000 cases?

The favorability of the AJCA regime depends on our view of the counterfactual. We observe a relative increase in the proportion of sub-$440,000 cases. The worst situation is that the relative increase is due solely to the relative decrease in higher value cases in the AJCA regime. If the AJCA’s effective increase in bounties somehow counter-intuitively caused whistleblowers to participate less in high value cases, this would make the change undesirable. On the other hand, if the increase in bounties actually deterred high-value fraud from occurring in the first place, the law would seem to be a success.

Assuming that the increase in sub-$440,000 cases is due to an absolute volume increase in comparison to the pre-AJCA regime, there are at least three possible counter-factual conditions for these sub-$440,000 cases. First, there are non-marginal cases that would still have been prosecuted under the pre-AJCA regime and were of the same value. For these cases, the increased bounty was detrimental to government compensation. The DOJ could have paid less to the relator and still have prosecuted the case.

The second category consists of cases that would not have been prosecuted under the pre-AJCA regime. These cases generate new compensation that would not have otherwise been available to the government. Their prosecution may also generate other deterrence or social benefits, as observers may feel that crime is less likely to pay. Prosecution of these additional cases is not costless, though. The DOJ will face additional burdens due to these filings, as may the judiciary. The overall desirability of this increased enforcement depends on the difficult cost-benefit comparison.

The third category of cases is the most desirable. These are cases that would also have been prosecuted under the pre-AJCA regime but for much greater losses. In the pre-AJCA regime, the fraud might have accumulated to $1,000,000 before prosecution; the increased AJCA bounty resulted in prosecution before the fraud reached $1,000,000 in losses. The earlier prosecution could be a result of the whistleblower or her attorney acting sooner because her recovery share was sufficient to offset the cost of participating. The earlier prosecution could also be the result of greater investments by relators or attorneys in detecting these lower value frauds. Regardless of the specific mechanism,
this category of cases does not impose any additional burdens on the DOJ or the judiciary. The government would have handled the case regardless of the bounty increase. I describe this effect as “incapacitation,” as the defendant was unable to commit the full extent of fraud desired prior to prosecution.

The composition of the sub-$440,000 cases under the AJCA is likely a mix of the aforementioned categories. I do not yet have a strategy for identifying the mix of cases, but the greater the proportion of the third category, the more desirable the bounty increase. Utilizing an increased bounty percentage to catch cases earlier is administratively and judicially attractive in comparison to complex proposals looking at “constructive knowledge.”\footnote{See Kovacic, supra note 23.} The simpler bounty increase may also incur less strategic responses to added legislative complexity.

Weaknesses of this approach

The empirical strategy I have chosen has quite a few limitations. First, similar to regression discontinuity designs, the window of causal inference is extremely narrow, which may limit the external validity. Even if everything in the measurement design works well, the strongest claim I can make is that, given the bounty situation that existed before October 2004, an approximate 23% post-tax increase in bounty generated an increase in relator participation for cases valued under $440,000. This claim is, of course, different than saying that an additional 23% today would still generate a further increase in relator participation. This difficulty in causal inference is unfortunately common, but I believe these data at least give policy makers some evidence as to future considerations in the finder’s bounty percentage.

Second, there are legitimate concerns with the difference approach I utilize instead of the preferred regression discontinuity design. Given the likelihood of censored data in the last five years of my data and the limited number of cases after 2004, the difference approach seems to be my best choice at present. Nonetheless, since I am not comparing the absolute volume of cases before and after October 2004, the design does not actually show an absolute increase in volume of cases valued under $440,000. The increase is relative to higher valued cases. As such, from an absolute sense, there may not have been an actual increase in case volume after October 2004 attributable to the AJCA. As more cases filed after 2004 are resolved, we may be able to gain better insight into the absolute effects of the AJCA.

Third, the first stage of my exogenous variation could use more support. It is unclear how attorneys or relators perceive a 23% increase in the effective bounty percentage. The percentage does not seem so large as to “shock” marginal individuals into taking action. If relators and attorneys do not have clear conceptions of the value of their cases, it is less certain if a 23% increase would be meaningful.

Fourth, the censoring effect makes it difficult to determine the general quality of cases submitted to the DOJ after the change in bounty. It is entirely possible that an increased
bounty resulted in an increased volume of low quality cases. Submission of these low quality cases could be costly to many of the parties involved: the DOJ expending resources investigating a weak claim, the potential defendant expending resources in cooperating with an investigation or defending itself, and the relator suffering the repercussions of whistleblowing. Nonetheless, just because a case may be lower quality does not necessarily mean it has no value; it might, for example, inform the government of a previously unrecognized avenue fraud, even if the specific instance identified might not be worth prosecuting. Properly informed analysis would also weigh the costs of addressing the poor quality cases against the benefits from greater enforcement.

D. Conclusion

Private individuals have valuable information that may help prosecute cases of fraud, but the choice to become a whistleblower is not easy. By paying a finder’s bounty, the government can encourage these whistleblowers to come forward. Such a whistleblower reward program can help the government both recover penalties against offenders and potentially deter people from committing fraud in the first place.

One difficult question, then, is selecting the proper finder’s bounty percentage. If the bounty percentage is too low, perhaps too few whistleblowers will come forward. Conversely, an excessively high bounty percentage may leave “money on the table,” as all of the whistleblowers who participate would have been willing to do so for less money.

I utilize evidence from the False Claims Act to suggest that the Federal government’s base bounty of 15% for information regarding fraud may be too low for certain offenses. Given the spectrum of fraud offenses, whistleblowers seem to address only a limited number of cases valued under $440,000. Working with evidence from a 2004 change in the tax code, I claim that whistleblowers may be more willing to disclose information regarding these smaller offenses in exchange for a higher bounty percentage.