

**Carrots and Sticks of Whistleblowing:
What Classification Trees Say About False Claims Act Lawsuits**

*Tammy W. Cowart**

*Kurt S. Schulzke***

*Sherry Jackson****

Key Words: False Claims Act; Whistleblowing; Classification Tree; Legal Cases

ABSTRACT

Whistleblower lawsuits under the federal False Claims Act have markedly increased over the past decade. While the amount of individual settlements and judgments vary, over time the government has paid out an average of about 15% of its recoveries as awards to whistleblowers. This investigation used a classification tree algorithm to analyze a sample of recent False Claims Act *qui tam* settlements, identifying several factors that distinguish larger settlements from smaller ones. Notably, the public or private status of corporate defendants, the federal judicial circuit in which the case is settled, type of case, case duration, relator status, and whether the Department of Justice intervened in the case are significant indicators of settlement amount in *qui tam* cases. Companies can use this information to better evaluate their exposure to liability for whistleblower litigation, while potential whistleblowers and their counsel can use it to evaluate the likelihood of winning an award.

INTRODUCTION

According to a 2013 study by the Ethics Resource Center, 63% of employees who observe misconduct on the job will report it, up from 58% in 2007.¹ In addition, 21% of those employees complain of retaliation after reporting misconduct, up from only 12% in 2007.² In a separate survey of Fortune 1000 public relations executives, 33% of respondents who had reported wrongdoing reported suffering some form of retaliation.³ Thus, both whistleblowing and complaints of retaliation against whistleblowers are rising. Whistleblowers face enormous pressures in deciding whether to pursue a whistleblower case. The prospect of retaliation can be daunting, while the potential for a whistleblower award can provide some incentive to ‘do the

*J.D., Associate Professor of Business Law, University of Texas at Tyler

**J.D., C.P.A., C.F.E., Associate Professor of Accounting & Law, University of North Georgia

***Ph.D., Associate Professor of Management, University of Texas at Tyler

¹ Ethics Resource Center, *National Business Ethics Survey*, Washington, D.C. (2014) (last visited January 4, 2018).

² *Id.*

³ C. A. Greenwood, *Whistleblowing in the Fortune 1000: What practitioners told us about wrongdoing in corporations in a pilot study*, 41 PUB. REL. REV. 490 (2015).

right thing.’ Cho and Song describe this ethical dilemma as the effect of individual perceptions and motivation on whistleblower behavior.⁴

Cases brought by whistleblowers can also have a significant impact on a company. For example, financial penalties and related sanctions—such as compensation for retaliation, relator’s attorney fees, and costly corporate integrity agreements—may be imposed under the False Claims Act (FCA) or through whistleblower award programs administered by the SEC, CFTC, and IRS. These can also affect consumer confidence and investor satisfaction.⁵ The loss of corporate reputation can also be damaging, so companies are well-advised to understand the factors that impact large whistleblower-related payouts. The False Claims Act is currently one of the most active whistleblower rubrics under U.S. law and is the subject of this study.

Under the False Claims Act (FCA), a private whistleblower called a “relator” can file a lawsuit in federal district court alleging that a non-government actor or municipality fraudulently or recklessly misappropriated government funds through so-called “false claims”.⁶ Such lawsuits are called “*qui tam*” claims because the relator sues for the benefit of both the government and the relator. The complete Latin phrase is “*qui tam pro domino rege quam pro se ipso in hac parte sequitur*,” meaning in essence, “he who sues for the king and for himself”.⁷ FCA cases can also be initiated by the government acting on its own behalf. It is useful to keep in mind that an FCA claim does *not* require an allegation of fraud; recklessness suffices.⁸

In terms of the volume of new cases filed, whistleblower activity under the FCA is at an all-time high. In total, *qui tam* filings over the most recent ten years, an interval in which federal budget outlays increased by more than 70% (OMB), are up 37% over the previous ten years.⁹ Nationwide, a total of 6,494 new FCA cases were opened in the 2005-2014 interval, of which 5,298 (81.6%) were relator-initiated *qui tams*.¹⁰ In the preceding decade, 1995-2004, 74.9% of 5,177 new FCA cases were *qui tams* (DOJ, 2014).¹¹

Settlements, however, have not kept pace with new filings. An unofficial tally of FCA settlements and judgments between 2000 and 2015 indicates that 1,369 FCA cases were resolved in the 2004-2014 interval, with 1,277 resolved over the 10 years ending in 2014.¹² Based on the

⁴ Y. J. Cho & H. J. Song, *Determinants of Whistleblowing within Government Agencies*, 44 PUB. PERSONNEL MGMT. 450 (2015).

⁵ D. Driscoll et al., *Business Ethics and Compliance: What Management is Doing and Why*, 99 BUS. & SOC’Y REV. 35 (1998).

⁶ False Claims Act, 31 U.S.C. §§ 3729-3733 (1994).

⁷ JAMES B. HELMER, FALSE CLAIMS ACT: WHISTLEBLOWER LITIGATION 69-70 (Bloomberg BNA, 6th ed., 2012).

⁸ *Id.*

⁹ CIVIL DIVISION, UNITED STATES DEPARTMENT OF JUSTICE, *Fraud Statistics - Overview, October 1, 1987 - September 30, 2014* (2014) <http://www.justice.gov/civil/pages/attachments/2014/11/21/fcastats.pdf> (last visited January 4, 2018).

¹⁰ See TAXPAYERS AGAINST FRAUD, FCA STATISTICS 2004-2014, <http://www.taf.org/DOJ-FCA-Statistics-2014.pdf> (last visited January 4, 2018).

¹¹ Civil Division, *supra* note 9.

¹² B. Elmer, *False Claims Act Settlements 2000-2015*. (Apr. 23, 2015), <https://www.crowell.com/files/False-Claims-Act-FCA-Settlements-Crowell-Moring.pdf>.

sample examined in this study, the median time interval between filing the original *qui tam* complaint and resolution is four years. However, ten years is not uncommon. Over the same period, the federal government recovered \$21.6 billion (excluding criminal fines and forfeitures) through *qui tam* FCA actions, paying out \$3.3 billion in relator awards¹³ or roughly one-hundredth of one percent of the \$31.9 trillion in budgeted federal government outlays (OMB) over the same period. Assuming the accuracy of Elmer’s count of resolved FCA cases, this translates to a mean of \$2.5 million per award. Thus, the government invested \$330 million per year in *qui tam* awards, yielding taxpayers an ROI of 654%. Over the same interval, without whistleblower help, the government recovered \$9.4 billion from its own FCA actions, at an undisclosed cost to taxpayers.¹⁴ Along similar lines, in the 2011-2015 interval, the DOJ recovered more than \$3.5 billion in settlements and judgments from civil FCA cases. Moreover, in fiscal year 2015 the government paid \$597 million to *qui tam* relators.¹⁵

For whistleblowers and experienced *qui tam* counsel, the economics and institutional dynamics of the FCA discourage the filing of even factually and legally solid cases alleging less than \$10 million in “single damages.”¹⁶ While the \$2.5 million per award might seem enticing, the reality is more complicated. First, the \$2.5 million is misleading because the distribution of resolution amounts on which the awards are based is skewed by a few outlier recoveries in excess of \$500 million. Second, a surprisingly small fraction of each recovery actually reaches the relator. If the government recovers \$10 million in a single case (a very big “if”), a solo relator’s typical 15% share starts at \$1.5 million and is quickly whittled down by contingent legal fees (e.g., 35-50%) and income taxes (roughly 45%, state and federal combined). After deducting 40% for legal fees and 45% in taxes on the remaining balance, only \$495,000 remains of the original \$1.5 million award. After tax, the *successful* relator’s attorneys would take home \$330,000 from the contingency, plus reasonable costs and reasonable hourly fees. The math in larger cases, which typically involve multiple relators and law firms who must share any award, is even more daunting.

Award dollars that remain in play at the individual level must be further reduced probabilistically and then discounted to present value over the expected duration of the case, taking into consideration the real possibility that a settlement or judgment may be severely reduced because of “ability to pay” issues or evaporate altogether in the defendant’s bankruptcy. The vast majority of whistleblower claims are eventually dismissed with zero recovery, most recoveries are relatively small, most whistleblowers lose their jobs and remain unemployed for years, and many are financially and emotionally devastated before an award, if any, reaches them.¹⁷ Attorneys for an unsuccessful relator—meaning one of the majority whose cases end in

¹³ Civil Division, *supra* note 9.

¹⁴ Civil Division, *supra* note 9.

¹⁵ OFFICE OF PUBLIC AFFAIRS. DEPARTMENT OF JUSTICE. *Justice Department Recovers over \$3.5 Billion from False Claims Act Cases in Fiscal Year 2015*. (2015) <http://www.justice.gov/opa/pr/justice-department-recovers-over-35-billion-false-claims-act-cases-fiscal-year-2015>.

¹⁶ In *qui tam* jargon, “single damages” refers to the actual amount of false claims (meaning what the government actually lost as a result of the false claims) before statutory doubling or trebling and per-claim fines.

¹⁷ C. C. Verschoor, *We Need More Whistleblowers*, 91 STRATEGIC FIN. 15 (2010).

dismissal with no settlement—will typically be “on the hook” for upwards of \$100,000 in costs for discovery, expert witnesses, and travel expenses.

Meanwhile, the federal government’s Centers for Medicare and Medicaid Services (CMS) itself estimates that 10% of its own \$603 billion in *annual* Medicare expenditures are “improper”.¹⁸ Malcolm Sparrow testified in 2009 that the government’s 10% loss estimate is “sadly lacking in rigor” and “quite misleading,” that actual losses might be as high as 30%, and that even 10% is exponentially higher than the credit card industry’s fraud loss benchmark of one-tenth of one percent.¹⁹ Additional losses accrue at the state level and in federal non-Medicare programs such as defense, education, environmental protection, homeland security, and research grants. Given such towering fraud losses and the outsized taxpayer ROI on investments in relator awards, companies should not be too surprised if the government encourages more relators to come forward.

A good first step is to understand the factors that influence FCA payout amounts because gross FCA payouts are the mathematical basis for awards paid to relators. For the purpose of investigating these payout factors, we propose the use of a classification tree model. Ideally, the model would include factors such as the federal circuit in which settlement occurs, the FCA success records of various U.S. Attorney’s Offices (USAO) that investigate FCA allegations, the type of FCA case involved (Medicare/Medicaid; Anti-kickback, etc.), local and national political considerations relating to the sensitivity and timing of the case. In practice, however, much of the data relating to these variables is inaccessible because of attorney-client privilege issues and DOJ disclosure policies. Based on publicly available data, our model considers a variety of factors that conventional wisdom suggests should impact recovery amounts.

Information regarding the factors that impact the government’s recovery in FCA cases can help companies, government, relators, potential relators, and their respective legal counsel in making decisions like, for example, whether to file a claim at all, whether some candidate filing venues are better than others, and how best to respond to relator claims. The purpose of this paper is to identify and rank the most significant factors affecting the magnitude of government recoveries in these *qui tam* cases. In Part II, we will review the relevant law in whistleblower cases. Part III will discuss classification trees, and Part IV explains our methods. In Part V, we discuss the results of our classification tree analysis.

LITERATURE REVIEW

Whistleblower Legal Background

False Claims Act. *Qui tam* is hardly a new concept. The legal pedigree of *qui tam* lawsuits in England can be traced as far back as the early 14th century, during the reign of

¹⁸ GOVERNMENT ACCOUNTABILITY OFFICE. *Report to Congressional Committees, High-Risk Series Update*, p. 359. Washington, DC: U.S. Government Printing Office (2015), available at <http://www.gao.gov/assets/670/668415.pdf>.

¹⁹ *Criminal Prosecution as a Deterrent to Health Care Fraud: Hearing before the Senate Committee on the Judiciary, Subcommittee on Crime and Drugs*, May 20, 2009. (statement of Malcolm Sparrow), available at <https://www.gpo.gov/fdsys/pkg/CHRG-111shrg55465/html/CHRG-111shrg55465.htm>.

Edward III.²⁰ In the United States, the FCA was first passed in 1863 at the urging of President Abraham Lincoln who sought to combat rampant defense contract fraud against the Union Army during the Civil War.²¹ Since that time, the FCA has been known as “Lincoln’s Law”.²²

Eighty years later, Congress essentially gutted the FCA when, in the middle of World War II, relators themselves began raiding the government’s coffers by piggy-backing their own *qui tam* FCA claims on DOJ criminal investigations of wartime fraud. Congress’ undebated assumption was that the DOJ certainly must be capable of handling this kind of fraud without outside help.²³ Among the 1943 FCA amendments, Congress reset relator’s award ceilings at 10% and 25% for intervened and declined cases, respectively, but left the floors to the court’s discretion; in other words, “successful” relators could get nothing at all for their pains.²⁴ Not surprisingly, for about forty years thereafter the *qui tam* bar essentially disappeared and so did their relator clients.²⁵

By 1986, the data clearly showed that the DOJ was incapable of policing rapidly growing defense contract fraud on its own.²⁶ In response, Congress and President Ronald Reagan reenergized the FCA through amendments that (1) empowered the relator to continue as a party to a FCA civil action even if the DOJ intervenes, (2) reset and guaranteed relator’s shares at 25-30% and 15-25%, respectively, for declined and intervened cases,²⁷ (3) added anti-retaliation provisions, making it possible for relators to bring suit against employers who “retaliate, demote, or otherwise discriminate against an employee” who brings suit,²⁸ and (4) eliminated the 1943 “prior government knowledge” jurisdictional bar.²⁹ These changes, together with dramatic increases in federal government spending, have since resulted in a major increase in the volume of *qui tam* suits under the FCA. As referenced earlier, over 80% of all FCA suits are *qui tam* suits meaning that they are initiated by non-government relators.

Other Whistleblower Laws. Beyond the FCA are other federal statutes that harness whistleblowers for various regulatory purposes. The 2002 Sarbanes-Oxley Act included measures to protect SEC whistleblowers from retaliation. The Deficit Reduction Act of 2005

²⁰ HELMER, *supra* note 7.

²¹ *Id.*

²² B. Joshpe, *Celebrating the 150th birthday of 'Lincoln's Law': Privatized fraud fighting*, FORBES, Mar. 6, 2013, <http://www.forbes.com/sites/realspin/2013/03/06/celebrating-the-150th-birthday-of-lincolns-law-privatized-fraud-fighting/>.

²³ HELMER, *supra* note 7.

²⁴ HELMER, *supra* note 7.

²⁵ *Id.*

²⁶ *Id.*

²⁷ *Id.*; C. Broderick, *Provisions and the public interest: An empirical analysis*, 107 COLUMBIA L. REV. 949 (2007).

²⁸ G. Rapp, *Four signal moments in whistleblower law: 1983-2013*, 30 HOFSTRA LAB. & EMP. L.J. 389 (2013).

²⁹ HELMER, *supra* note 7.

incentivizes states to adopt statutes that mirror the federal FCA.³⁰ In the securities and financial reporting arena, the 2010 Dodd-Frank Act created FCA-like whistleblower reward and protection regimes for SEC and CFTC whistleblowers, partially duplicating Sarbanes-Oxley's SEC anti-retaliation rubric.³¹ In this vein, the SEC recently moved to restrict the use of employer-employee confidentiality agreements that could be construed as discouraging whistleblowers from reporting wrongdoing.³² Additional federal and state regulations endeavor to protect whistleblowers in various ways.³³ While the plethora of whistleblower laws might seem to dramatically increase risks for firms and virtually guarantee a favorable outcome for any whistleblower, the reality is that whistleblowing continues to be fraught with financial, professional, and reputational risk.

FCA Procedures. When a relator decides to bring a *qui tam* suit, he or she must be represented by an attorney. The complaint must be filed under seal and remain sealed for at least sixty days. At the conclusion of this period, the Department of Justice (DOJ) must file a motion to keep the case under seal if, as typically occurs, it chooses to do so.

While the relator must disclose the facts of the case to the Attorney General (AG), the government has no authority to prevent the filing or to force the relator to file in a particular federal judicial district. Furthermore, by statute, the DOJ must investigate all FCA *qui tam* claims that comply with minimum procedural requirements. However, relators customarily do not actively participate in case investigation after filing the complaint except as requested by the DOJ. Thus, the DOJ wields enormous power over case outcomes through allocation of investigative resources and discretion to intervene or decline the case. A relator who ignores DOJ preferences may find her case either back-burnered or declined.

While the case is under seal, the Attorney General must investigate the allegations in the complaint.³⁴ The seal initially lasts for 60 days but is typically extended—often for two years or more—to enable the DOJ to secretly investigate the claims.³⁵ As a practical matter, however, defendants often learn of the investigation before the seal is lifted because investigative maneuvers can be difficult to hide.

At the conclusion of the investigation, the DOJ can choose one of the following four options:

1. Intervene in one or more counts in the *qui tam* action, opting to lead the prosecution of the case.³⁶

³⁰ Rapp, *supra* note 28.

³¹ *Id.*

³² M. Boxer & N. Wang, *SEC restricts ability of companies to silence employees with confidentiality agreements*. 41 EMP. REL. L. J. 45 (2015).

³³ R. Moberly, *Sarbanes-Oxley's whistleblower provisions: Ten years later*. 64 S. CAROLINA L. REV. 2 (2012).

³⁴ 31 U.S.C. § 3730(a) (1994).

³⁵ 31 U.S.C. 3730(b) (1994). See The Process of False Claims Act Litigation, TAF EDUCATION FUND (last visited Jan. 13, 2018), https://www.taf.org/Public/Resources_by_Topic/FAC__False_Claims_Act/Processes/Public/Resources_by_Topic/FA__False_Claims_Act/Process.aspx (a seal period of two years or more is “not unusual”).

³⁶ 31 U.S.C. § 3730(c)(1) (1994).

2. Decline to intervene, leaving the relator and his or her attorney to prosecute the case on behalf of the United States.³⁷
3. Pursue a so-called alternate remedy such as an administrative proceeding in which the relator is entitled to the same award as if the matter were pursued under the FCA.³⁸
4. Move to dismiss the complaint.

Often, the DOJ's decision on intervention or declination is preceded by a "partial lift" of the seal that allows the DOJ to disclose the claim to the defendant and negotiate for a settlement. The vast majority of *qui tams* are declined because of resource constraints, opposition to the case by the allegedly victimized federal agency, or evidentiary or procedural weaknesses, as evidenced in part by declined cases that end in settlements or jury findings of liability. Declination, however, should not be viewed as the DOJ's judgment on the merits of the case. After declination, the government remains a party to the action and wields veto authority over any proposed settlement despite taking no active role in the case and bearing none of the costs of prosecuting it.³⁹

Of the declined cases, relators themselves dismiss most because of reluctance to risk scarce capital in litigation against well-funded defendants. Declined cases that are initially pursued by the relator are dismissed, settled, or—rarely—tried all the way to a jury verdict. While some declined cases end with large settlements—one company, threatened with loss of attorney client privilege, recently settled for \$450 million⁴⁰—some jury verdicts against companies are vacated after trial or overturned on appeal.⁴¹ The bottom line is that DOJ intervention is usually viewed by relators' counsel as a big victory, virtually clinching some award.

If a *qui tam* case survives through to settlement, trial, or alternate remedy, the relator is statutorily entitled to between 15 and 25% (in an intervened case) or 25 and 30% (in a declined case) of the civil damages and statutory fines recovered by the government.⁴² The FCA theoretically mandates treble damages (three times the government's actual losses from the false claims) and civil penalties of between \$5,500 and \$11,000 per false claim,⁴³ but the trebling may be reduced to doubling through the defendant's timely voluntary disclosure of the false claims.⁴⁴

³⁷ 31 U.S.C. § 3730(c)(3) (1994).

³⁸ 31 U.S.C. § 3730(c)(5) (1994).

³⁹ 31 U.S.C. § 3730(c)(1) (1994).

⁴⁰ U.S. ex rel. Vainer v. DaVita Inc. et al., 1:07-cv-02509, N.D. Ga (2015).

⁴¹ See U.S. ex rel. Absher v. Momence Meadows Nursing Center, Inc., 764 F.3d 699 (7th Cir. 2014); K. Goldberg, *6th Circ. Nixes \$83M Fresenius Medicare Fraud Judgment*, LAW360, Oct. 5, 2012, <http://www.law360.com/articles/384735/6th-circ-nixes-83m-fresenius-medicare-fraud-judgment>.

⁴² Relators who planned and initiated the fraud are entitled to a percentage less than 15, subject to the court's discretion (see, Helmer *supra* note 7 at 915).

⁴³ 31 U.S.C. §3729(a)(1) (1994).

⁴⁴ 31 U.S.C. §3729(a)(2) (1994).

Similarly, the civil penalty amount is subject to the court's discretion.⁴⁵ In practice, doubling, trebling and civil penalties typically materialize only in the cases decided by juries.⁴⁶ Thus, these damage multipliers serve primarily to induce defendants to settle before trial and do not reliably predict actual eventual FCA liability except in outlier cases.

According to DOJ statistics, the civil portion of all FCA settlements and judgments in the 2004-2014 interval was \$31.70 billion. Of this total, \$22.22 billion (70%) came from relator-initiated *qui tam* claims, delivering \$3.39 billion (15.26 percent of related *civil* recoveries⁴⁷) in awards to relators. Of *qui tam* recoveries, the DOJ data indicate that roughly \$813 million (about 3.7%) came from declined cases, with 96.3% coming from intervened cases. This purported allocation of recoveries between declined and intervened cases is misleading, however, because the DOJ also counts as "intervened" cases in which the DOJ intervenes after most of the real litigation risks have been taken and the work is done by relators and their counsel. The DOJ statistics do not (a) distinguish between such "nominal" interventions and "real" ones, (b) provide a count of cases yielding recoveries, or (c) reveal any case-level information.

Classification Tree

Classification trees, sometimes also called "decision trees" (for categorical or binary response variables) or "regression trees" (for continuous response variables), are simple and effective methods used in data mining to explain relationships between predictor variables and a target outcome.⁴⁸ Trees are also used as a preliminary data screening tool to discover meaningful patterns in large and complex data sets as a step toward fitting regression or other types of statistical models.⁴⁹ A classification tree maps inputs to predefined classes or binary outcomes, such as win or lose, or guilty or not guilty. Prior researchers suggest that classification trees are useful to predict the outcome of patent cases⁵⁰ and search and seizure cases.⁵¹ Classification trees

⁴⁵ Helmer *supra* note 7 at 849.

⁴⁶ See, e.g., U.S. ex rel. Drakeford v. Tuomey, 2015 U.S. App. LEXIS 11460 (4th Cir. 2015) (affirming a treble-damages-plus-civil-penalties judgment of \$237,454,195 against a regional healthcare system for knowing submission of 21,730 false claims to Medicare); Lisa Schenker, *\$237 Million Tuomey Judgment upheld by Federal Appeals Court*, MOD. HEALTHCARE, July 2, 2015, <http://www.modernhealthcare.com/article/20150702/NEWS/150709975> (providing additional background on the Tuomey case).

⁴⁷ Relators are statutorily entitled to a share of government recoveries obtained through "alternate remedies" such as related criminal fines, penalties, or forfeitures (Helmer, *supra* note 7) but DOJ statistics do not report amounts recovered through such alternate remedies.

⁴⁸ Mira Shapiro, *Using JMP® Partition to Grow Decision Trees in Base SAS®*, (2013), <http://analytics.ncsu.edu/sesug/2013/JMP-04.pdf>; Russ Lavery, *An Animated Guide: Regression Trees in JMP® & SAS® Enterprise Miner™*, (2012), <http://www.lexjansen.com/nesug/nesug12/sa/sa05.pdf>; Lior Rokach & O. Maimon, *Data Mining with Decision Trees* (World Scientific, Singapore 2007).

⁴⁹ Lavery *supra* note 48 at 6.

⁵⁰ Tammy Cowart, et al., *Two Methodologies for Predicting Patent Litigation Outcomes: Logistic Regression versus Classification Trees*, 51 Am. Bus. L. J. 843 (2014).

have been used in a variety of settings, including finance, marketing, engineering, medical diagnosis, quality control, and credit evaluation.⁵²

Many variables can be considered in decision making. Classification trees can be used to help identify independent variables and interactions that have the strongest statistical association with the classification or decision outcome.⁵³ Trees provide a useful, visual model that explains how a decision could be made.⁵⁴ Classification trees consist of nodes where data are partitioned, based on the value of an input characteristic, into one or more categories with similar values and, therefore, approximately the same probability of attaining some common outcome. The classification tree is created when the data set is loaded into the classification tree software, which has identified variables and values. The software evaluates the data, classifies it based on the variables, and produces the visual classification tree. The tree provides the visual framework that specifies the sequence of variables and values that are associated with the classification or decision process.⁵⁵

Classification trees are usually represented graphically as a hierarchical structure.⁵⁶ The classification tree begins with the trunk, which then branches out based on input variables and their interactions to create branch patterns. Graphically the tree can be presented from top to bottom or from left to right.⁵⁷ The graphical presentation makes the results self-explanatory and easy to interpret.⁵⁸ This is especially beneficial in a practitioner setting⁵⁹ such as medical diagnosis or legal triage.⁶⁰ In this study we use classification trees to identify key variables and

⁵¹ Jonathan Kastellac, 7 J. *Empirical Legal Stud.* 202 (2010).

⁵² P. Austin, *A comparison of regression trees, logistic regression, generalized additive models, and multivariate adaptive regression splines for predicting AMI mortality*, 2 *Stat. in Med.* 2937 (2007); R. Guh & Y. Shiue, *On-line identification of control chart patterns using self-organizing approaches*, 43 *Int'l J. of Production Res.* 1225 (2005); B. Rosenfield & C. Lewis, *Assessing violence risk in stalking cases: A regression tree approach*, 29 *LAW & HUM. BEHAV.* 343 (2005); S. Rayo & A. Cortes, *Applying chaid to identify the accounting-financial characteristics of the most profitable real estate companies in Spain*, 15 *J. OF ECON., FIN. & ADMIN. SCI.* 51; C. Wang, et al., *Decision tree based control chart pattern recognition*, 46 *INT'L J. OF PRODUCTION RES.* 4889 (2007).

⁵³ T. MITCHELL, *MACHINE LEARNING*, (McGraw-Hill 1997).

⁵⁴ S. ERIKSEN & R. KELLER, *DECISION TREES*, IN *ENCYCLOPEDIA OF OPERATIONS RESEARCH AND MANAGEMENT SCIENCE*, 139 (Saul I. Gass & Carl M. Harris eds., 1996).

⁵⁵ ROKACH & MAIMON, *supra* note 48.

⁵⁶ *Id.*

⁵⁷ S. ERIKSEN & R. KELLER, *supra* note 54.

⁵⁸ ROKACH & MAIMON, *supra* note 48.

⁵⁹ M. Tonkin et al., *A comparison of logistics regression and classification tree analysis for behavioral case linkage*, 9 *J. OF INVESTIGATIVE PSYCHOL. & OFFENDER PROFILING* 235 (2012).

⁶⁰ B. Rosenfield & C. Lewis, *Assessing violence risk in stalking cases: A regression tree approach*, 29 *LAW & HUMAN BEHAV.* 343 (2005).

their interactions that impact the government's civil recovery amounts in FCA cases, predefining the outcome variable as small (less than \$10 million) or large (greater than \$10 million).

METHOD

Sample Selection and Variable Description

While the DOJ reports annual aggregate settlement data,⁶¹ because of the unique procedure, politics, and institutional dynamics of the FCA, there is no official repository of data on FCA *qui tam* recoveries at the individual settlement level. As a result, data acquisition for this study began by selecting a non-random subset of the "major" FCA *qui tam* case resolutions (settlements or verdicts) reported by Taxpayers Against Fraud, Inc. (TAF) as occurring between 2004 and 2014. For these resolutions, some of the independent variables could not be derived from the TAF data. Therefore, necessary data were extracted from a variety of sources including press releases, news stories, court filings, law firm websites, and the DOJ's website.

During this second stage of data acquisition, the initial table of resolutions was refined by adding some observations discovered during the research and eliminating others for which key variables were missing. For example, TAF might identify a single resolution of \$100 million which, upon closer examination, was the sum of two separate resolutions associated with related but distinct cases. The final data table comprises 206 resolutions, the descriptive statistics for which appear in Table 2 *Appendix B*.

Government FCA Recovery Dependent Variable.

The dependent variable is binary, indicating whether the government's recovery was less than or greater than 10 million dollars. We selected 10 million dollars as a cutoff because at this amount, there was a relatively even percentage split in the data. Please see Table 1 for a complete description of the outcome variable. *Appendix B*

Independent Variables

Six independent variables were used to develop the classification tree. Table 2 provides detail on the variables and the breakdown of recoveries by variable codes. The following sections describe the six independent variables. *Appendix C*

Duration. FCA cases are notoriously long-lived. Duration was measured as the time interval in whole years between case filing and case settlement. Duration data were available for just over half (109) of the cases in the sample. Among these cases, 27 ran on for more than five years and four went beyond 10 years. From a theoretical prior expectation standpoint, longer duration might logically be associated with higher expected settlement value because relators and their attorneys are unlikely to continue fighting for a case without the expectation of a large

⁶¹ Civil Division, *supra* note 9.

eventual payout. However, viewed prospectively, the longer it is expected to take to bring in a settlement, the lower the present value of the case to government, relator, and relator's counsel.

Relator. The Relator variable reflects the functional role or organizational status of the relator whistleblower. Relators were coded categorically as Emp (employee of the defendant), Mgr (manager of the defendant), Patient (medical patient), Org (organization such as another corporation), Industry Insider (usually a competitor of the defendant), Other, and Govt (a non-relator government agency). In theory, employees, patients, and managers should possess more valuable case-related information than others and FCA cases brought by them should be expected to settle for larger amounts.

Public or Private. Public or Private refers to whether the defendant was a publicly traded company (Public) and, therefore, subject to SEC registration and reporting requirements or not publicly traded (Private) and, therefore, not subject to SEC registration and reporting. The public or private status of corporate defendants (as opposed to individual defendants) is in part an indicator of company size; public companies tend to be larger than private ones. Therefore, public companies, on average, should be financially capable of paying large FCA settlements than are private companies.

DOJ Intervention. There were two codes for this variable: (1) Yes, indicating that DOJ did intervene and (2) No, indicating that DOJ did not intervene. DOJ intervention is possible only in *qui tam* cases; in FCA cases originated by the DOJ itself, intervention is neither necessary nor possible.

Circuit. The Circuit variable reports the federal judicial circuit in which, at the time of settlement or judgment, the court with jurisdiction over the case was located. Twelve judicial circuits are represented in the sample: First through Eleventh, plus District of Columbia. The circuit can impact the outcome of a case in part because the circuits differ in how they apply and interpret the FCA and in their related procedural rules. For example, the circuits currently disagree over the interpretation of Federal Rules of Civil Procedure⁶² (FRCP) 9(b) which, despite the fact that FCA cases are technically not fraud cases⁶³, has been held by all federal circuits to require FCA complaints to plead fraud "with particularity".⁶⁴ The Fourth, Sixth, Eighth, and Eleventh Circuits apply this particularity requirement most stringently, while the First, Fifth, Seventh, and Ninth Circuits less so.⁶⁵ Thus, cases venued in the former circuits should have correspondingly lower expected settlement values than those in the latter circuits because those in the former are more likely to be dismissed before the discovery process begins.

Case Type. The data table included nine case type categories: (1) defense, (2) pharmaceutical, (3) university grant, (4) Medicare/Medicaid, (5) FLSA, (6) AKS, (7) mortgage, (8) hospice/nursing home, and (9) other. Defense cases (1) allege fraud in defense contracting. Pharmaceutical cases (2) allege fraud (except violations of the anti-kickback statute) by pharmacies or pharmaceutical manufacturers. Often, such fraud involves so-called "off-label

⁶² FED R. CIV. P. 9(b).

⁶³ Helmer *supra* note 7 at p. 576.

⁶⁴ G. B. Breen, et al., *Supreme Court declines to opine on circuit split over Rule 9(b) pleading requirements for FCA claims*. EPSTEIN BECKER GREEN HEALTH CARE & LIFE SCIENCES CLIENT ALERT (Jun. 5, 2014), <http://www.ebglaw.com/news/supreme-court-declines-to-opine-on-circuit-split-over-rule-9b-pleading-requirements-for-fca-claims/>; Helmer *supra* note 7 at p. 577.

⁶⁵ *Id.*

marketing” of prescription medications written for purposes not approved by the FDA. University grant cases (3) allege failures by universities or other grant recipients to comply with grant requirements. Medicare/Medicaid cases (4) allege false claims on Medicare or Medicaid other than those involving pharmaceuticals, AKS, or hospices and nursing homes. FLSA cases⁶⁶ (5) allege wage and hour-type violations of the Fair Labor Standards Act. AKS cases (6) allege claims made on Medicare or Medicaid for medical goods or services (whether or not otherwise necessary and legitimate) provided to patients recruited through kickbacks paid in violation of the federal Anti-Kickback Statute.⁶⁷ Mortgage cases (7) typically allege fraud in relation to federally-subsidized home mortgages. Hospice/nursing home cases (8) allege overcharging, failure to provide required services or medication to patients, or admission to hospice care of patients who are not terminally ill. Category (9) is the catch-all for case types whose frequency did not appear sufficient to justify a separate category.

Classification Tree

We used Classification and Regression Tree (CART) software to analyze the association between the independent variables and the dependent variable and develop the classification tree.⁶⁸ The CART software is based on the original CART code developed by Stanford University and University of California at Berkeley statisticians Breiman, Friedman, Olshen and Stone. The CART algorithm automatically searches for important patterns and relationships present in the data.⁶⁹ CART is one of several common algorithms for the creation of classification trees. CART constructs only binary trees, which results in each internal node having exactly two outgoing “edges” or directed lines. A variable or “subtree” composed of multiple variables can appear in more than one branch of the tree.⁷⁰

Results of Classification Tree

Figure 1 shows the classification tree produced by the CART algorithm. The tree includes only the independent variables, presented in Figure 1 from top to bottom, found by CART to have a statistically significant association with the dependent variable (Recoveries)... At each node, the tree shows the number and percentage of cases with small (<\$10 million) and large (>\$10 million) recoveries. (*Appendix D*)

The classification tree resembles an upside down, physical tree and thus is read from top to bottom. Boxes on the tree are called “nodes.” The top node is the “root” and each node at the

⁶⁶ The data set used for this paper includes only one FLSA case, which the CART software automatically grouped with Defense cases in building the classification tree. This means that the FLSA case does not stand on its own as a single observation representing a separate category. The authors believed it inadvisable to drop this data point from the data set.

⁶⁷ 42 USC §1320a-7b(b), commonly referred to as “AKS.” A summary of the AKS and related Stark Law is available at <http://oig.hhs.gov/compliance/provider-compliance-training/files/StarkandAKSChartHandout508.pdf>.

⁶⁸ CART® Classification and Regression Trees. Basic SPM v7.0. (2013). *Computer Software*. Salford Systems, available at <http://www.salford-systems.com/products/cart>.

⁶⁹ D. STEINBER & P. GOLLA, CART 6.0 USER MANUAL (Salford Systems, San Diego, CA 2006).

⁷⁰ See ROKACH & MAIMON, *supra* note 48; Pang-Ning Tan, Michael Steinbach, and Vipin Kumar, INTRODUCTION TO DATA MINING 170 (2005).

end of a branch is a “terminal” or “leaf” node. Remaining nodes are called “interior” or “intermediate” nodes. Each split partitions the observations at the node according to the values or “levels” of the independent variable at that split. For example, the left branch of the tree splits on Case Type, Circuit, Relator, and DOJ Int. At the DOJ Int split, CART partitioned the eleven observations that reached this split between three “No” and eight “Yes” values of DOJ Int.

At each split, the node on the right represents a subset of cases with a higher frequency of large payouts (> \$10 million) than the node on the left. In the classification tree shown in Figure 1, the first and, therefore, statistically most important variable was the defendant’s publicly- or privately-held status. All public companies split to the right while private companies split to the left. The classification tree indicates that cases against non-publicly-held defendants paid out lower amounts more frequently (71% of cases) and while cases against publicly-traded defendants paid higher payouts more frequently (72% of cases). Case type was a key predictor for both public and private entities. For public companies, case types of Pharmaceutical, Medicare/Medicaid, AKS, Mortgage, and Hospice/Nursing had the highest probability of a higher payout (84%). While the opposite node shows that cases involving Defense or FLSA had only a 39% likelihood of paying a large settlement, where the relator was a manager, industry insider, or government agency, 75% of the settlement amounts were in excess of \$10 million. This suggests that the status of the relator in Defense or FLSA cases has a significant impact on the amount of the settlement.

Following the tree to the left we see that private entities with case types of university grants or Medicare/Medicaid, filed in judicial circuits 1, 2, 4-10, and any relator other than an organization had the highest likelihood of a small payout (89%). However, the classification tree does indicate that private company defendants in intervened Medicare or University grant cases venued in circuits 3, 11, or 12 paid settlement amounts in excess of \$10 million with a frequency of 87%. This interaction in the classification tree represents the highest risk of large settlements among private companies.

Table 3 lists the variables by relative importance. Case type had the greatest impact on the payout amount, followed by the Public or Private Entity status of the defendant, the Federal Circuit where the case was settled, the Duration of the case, Relator type, and, lastly, DOJ Intervention. The histogram in Table 3 is calculated based on the variable with the greatest impact, Case Type. The impact percentages of the remaining variables are based on the impact of Case Type on recovery amounts. *Appendix D*

An important part of the classification tree process is validating or assessing the quality of the underlying classification model. Prediction accuracy, or how well the classifier predicts the actual results, is a key model validation tool. To this end, Table 4 reports the prediction accuracy of this classification tree. Table 4 indicates that the classification tree model correctly predicts 74% of settlements greater than \$10 million and 67% of settlements less than \$10 million, for an overall prediction success rate of 70%. The prediction rate is somewhat analogous to the R^2 statistic in linear regression, which represents how well the model fits the data, while the overall prediction rate in classification tree analysis represents how well the tree framework predicts the outcome. Just as in linear regression, while it is theoretically possible to achieve a 100% prediction rate (model fit), a model that is a “pure” fit for the training data would be “overfitted” and thus have little predictive usefulness in relation to new data.⁷¹ In this study, the

⁷¹ Lavery *supra* note 48.

classification tree performs better at predicting larger payouts than smaller ones. The overall prediction rate of 70% is fairly robust and can be useful to companies and practitioners in evaluating potential outcomes of FCA cases. *Appendix E*

DISCUSSION

Application of classification trees

When a company discovers that it is the subject of an FCA investigation, the most common reaction is to immediately call outside legal counsel to manage the company's response to the government's investigation. Typical first steps also include preserving potentially relevant electronic and hard copy documents, interviewing employees or third parties with information about the complaint, and communicating with government investigators, management, and the board of directors.⁷² The precise strategy followed by the defendant should be informed by an understanding of factors known to be statistically associated with litigation outcomes. Classification trees are a good place to start. The information that can be gathered from this classification tree analysis can be useful for companies and practitioners. For example, public companies have far greater potential for liability than private companies. As mentioned above, case types of Pharmaceutical, Medicare/Medicaid, AKS, Mortgage, and Hospice/Nursing had the highest probability of a higher payout. Where the suit is filed is also influential, with 2nd, 6th, and 9th circuits being the most common places to file. Logically, if the DOJ intervenes in the case, there is also a greater likelihood for a settlement, and the data supports this as well.

Since use of classification trees in academic studies is a fairly new development, it might be helpful to illustrate the application of the classification tree model to specific cases to demonstrate the model's practical usefulness. We will examine two FCA claims that were included in the sample.

The first, a Medicare/Medicaid fraud *qui tam* claim against Arthritis and Allergy Associates, settled in 2010 for \$227,895. From this total recovery or settlement amount, the relator received an award of \$41,000 or 18%. In the classification tree, the first factor associated with the eventual settlement size is the privately held or publicly traded status of the defendant. Thus, the Arthritis and Allergy Associates case begins at the top or root node and then moves to the left side of the tree, following the Private (meaning not publicly traded) branch. At Case Type, the next factor, this case again moves left into the "Univ, Mcare" node. The case was settled while under the jurisdiction of the federal district court in Connecticut, which is located in the 2nd Circuit, so the case again moves left into the "All Other" node. Finally, at the Relator split, the case moves to the right into the "All Other" leaf node. Cases ending in this node produce small settlements (less than \$10 million) with a relative frequency of 89%.

The second case involves a large public company, Hewlett Packard. This *qui tam* suit for violation of the AKS was also resolved in 2010. Hewlett Packard paid \$55 million to resolve the case and the whistleblower received \$1.89 million. This case progresses through the classification tree beginning with the public company branch. Again, the case type is the next most important factor, so this case moved right into that node where the branch terminates in a leaf node. Cases ending in this node produced large settlements (more than \$10 million) with a relative frequency of 84%. Thus, we can say that for public company defendants in all types of

⁷² P.B. Murphy, et al., *How companies should respond to whistleblower complaints*, RISK MGMT. (Dec. 2, 2014), <http://www.rmmagazine.com/2014/12/02/when-the-whistle-blows-how-companies-should-respond-to-whistleblower-complaints/>.

cases other than Defense or FLSA, the only significant factor in differentiating large and small settlement amounts is the Case Type. By contrast, for public company defendants in Defense or FLSA cases, Relator type and (maybe) case duration also play a significant role.

Whistleblower prophylactics

Without a doubt, companies should prepare to defend themselves against whistleblower litigation by using classification trees and other tools, since they may also incur significant legal and other business-related costs.⁷³ For example, this classification tree analysis indicates that firms in the pharmaceutical industry are exposed to the highest probability of high absolute FCA payouts, among both privately held and publicly traded firms. Because of their greater vulnerability, pharmaceuticals companies would be well-advised to pay special attention to promoting an ethical, whistleblower-friendly environment. This can avert a more complex government investigation and negative public attention, which is better for the entire company.

In general, companies would be wise to implement policies that promote the internal reporting and early resolution of such concerns so that the company can limit financial and reputational damage by self-reporting to the government if warranted or, even better, by catching potential violations before they occur. As Dervan⁷⁴ suggests, the challenge for companies is to create an environment where employees feel secure reporting potential misconduct internally, that their concerns will be properly investigated and addressed, and that they will not be subjected to retaliation from bringing concerns forward. Similarly, MacGregor et al. propose that prerequisites of an effective whistleblower environment include (a) hiring the right people and properly developing them as willing whistleblowers, (b) adequately educating employees to recognize wrongdoing and to know how report it internally, (c) promoting reporting in the organization through rewards and institutional safeguards, and (d) ensuring appropriate analysis and resolution, including clear communication of investigative results to whistleblowers.⁷⁵ In addition, Peeples et al. suggest appointing a corporate ethics officer and providing multiple avenues for potential whistleblowers to report their concerns.⁷⁶

CONCLUSION

Classification tree analysis has the ability to graphically represent a decision making process that companies, government, whistleblowers, and their attorneys can use in assessing the factors that may influence the outcome of a FCA *qui tam* case. When wrongdoing is reported internally, companies can evaluate the potential liability involved and respond accordingly to the whistleblower's concerns. On the other side, prospective FCA whistleblowers can easily see that FCA payouts by publicly held defendants tend to be larger than payouts by others and that DOJ matters somewhat against private defendants but not as much as does the choice of judicial circuit.

⁷³ D. K. Peeples, et al., *When the Whistle is Blown*, 48 BUS. & SOC'Y 467 at 478-79 (2009).

⁷⁴ Lucian E. Dervan, *Responding to potential employee misconduct in the age of the whistleblower: Foreseeing and avoiding the hidden dangers*, 3 CORPORATE L. J. 670 at 670 (2008).

⁷⁵ J. MacGregor, et al., *Creating an Effective Whistleblowing Environment*. 96 STRAT. FIN. 35 (2014).

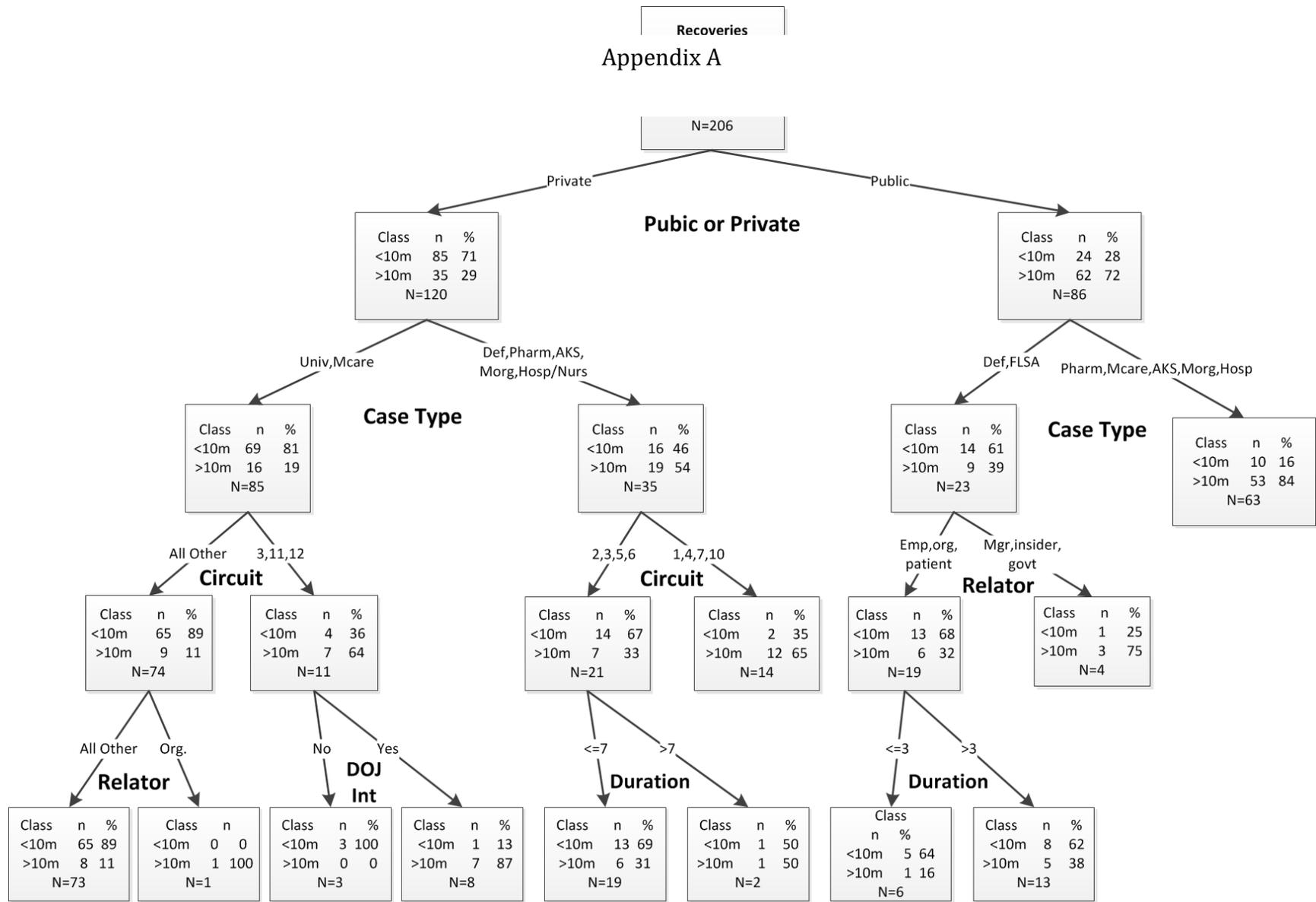
⁷⁶ Peeples, *supra* note 72 at 479-83.

A second advantage of the classification tree analysis is that it does not assume that the same predictor variable occurs in every case. The classification tree automatically evaluates interactions through its hierarchical and recursive process. For example, while the greatest exposure for publicly held companies lies outside of defense contracting and FLSA cases, defense contracting is on the high-risk side for private companies.

In this paper we examined various factors that impact the amount that the government has recovered in recent FCA whistleblower cases. We found that the public or private status of the company and case type had the biggest influence, followed by the federal circuit where the case was filed. These factor interactions can be useful to whistleblowers, companies, and their counsel, as well as to the government, in evaluating their respective prospects in future FCA litigation. With the increased inclusion of whistleblower protection and anti-retaliation statutes, companies cannot ignore the implications and potential risks of FCA litigation. Employers must be mindful of anti-retaliation statutes, establish effective processes to deal with concerns raised by employees and customers, properly train managers to respond appropriately to the concerns of potential and actual whistleblowers, and consult legal counsel when necessary.⁷⁷ Ultimately, as long as whistleblowers can voice their ethical concerns about company wrongdoing through the courts, it will be important to understand the forces that drive outcomes in whistleblower litigation.

Finally, we have suggested that effective corporate governance and sustainability require that the legal analysis facilitated by the classification tree be paired with high ethical standards and a whistleblower-friendly environment. We have pointed to evidence that companies will be better off in the long run if they foster ethical sustainability and encourage potential whistleblowers to bring their concerns forward early and internally to management. We have also highlighted specific steps that companies can take to promote these sustainability-enhancing objectives.

⁷⁷ Jamie D. Prenkert, *Handle with care: Avoiding and managing retaliation claims*, 55 BUS. HORIZONS 409 (2012).



Appendix B

Table 1 Dependent variable (recoveries)

Recoveries	N	%
< 10 million	109	52.9
> 10 million	97	47.1

Appendix C

Table 2 Descriptive statistics (independent variables)

Variables	Settlements <10 million	Settlements >10 million	Total
Federal Judicial Circuit			
1st	5	19	24
2nd	23	6	29
3rd	4	7	11
4th	5	7	12
5th	7	7	14
6th	11	6	17
7th	4	4	8
8th	3	3	6
9th	15	6	21
10th	6	4	10
11th	4	10	14
DC	5	4	9
Missing	17	14	31
DOJ Intervention			
Yes	60	81	141
No	49	16	65
Case Type			
Medicare/Medicaid	55	31	86
Pharma-Mkt	4	22	26
AKS	11	19	30
Mortgage	0	6	6
Hospice/Nursing Home	0	2	2
University Grant	7	1	8
Defense	13	8	21
FLSA	1	0	1
Other	16	9	25
Missing	2	0	2
Public or Private			
Public	24	62	86
Private	85	35	120

Duration			
0	0	3	3
1	5	3	8
2	11	7	18
3	5	6	11
4	10	9	19
5	8	9	17
6	2	11	13
7	1	6	7
8	1	5	6
9	1	2	3
11	0	1	1
12	1	0	1
15	0	1	1
17	0	1	1
Missing	64	33	97
Relator			
Manager	6	10	16
Employee	40	46	86
Organization	1	5	6
Patient	2	2	4
Industry Insider	5	11	16
Government	6	2	11
None	2	0	2
Missing	47	21	68

Appendix D

Table 3 Variable Relative Importance

Variable	Score %	Histogram
Case Type	100.0000	
Public or Private	87.8784	
Circuit	80.6695	
Duration	40.3877	
Relator	30.9221	
DOJ Intervention	15.8853	

Appendix E

Table 4 Prediction Accuracy

Actual	Total	% Correct
>10 million	109	74
<10 million	97	67
Overall	206	70
