



NOTE

Lies and Statistics: Statistical Sampling in Liability Determinations Under the False Claims Act

Patrick Kennedy*

Abstract. Medicare fraud costs this country billions of dollars a year and contributes to an ever-expanding debt. Conservatives want to cut spending on Medicare significantly, while liberals champion expanding Medicare. Finding common legislative ground between these positions has proven impossible. As a result, courts play an important role in pushing Medicare providers to stop defrauding the government. Unfortunately, up to the present, courts have rejected statistical methods of proof that could significantly reduce the cost of bringing suits against Medicare fraudsters. Judges who have ruled on the issue cite due process concerns with extrapolating that a care provider likely defrauded the government in a large number of cases from a much smaller subset of purportedly fraudulent claims. This Note provides guidance to courts, arguing that sampling is appropriate in cases against a single defendant with a large number of claims at issue where the alleged fraud is systematic and the variability between claims is relatively limited. Specifically, this Note deals with the realities of estimating a model that seeks to organize claims into categories of liability or nonliability.

* J.D. Candidate, Stanford Law School, 2019; Ph.D. Candidate (Economics), University of Illinois at Chicago, 2019. Thanks to David Studdert, Denise DeArmond, and Sophia Cai for their invaluable suggestions and editing help. Thanks as well to the editors of the *Stanford Law Review*, especially Isabella Sayyah, for their amazing editing and tireless work.

Table of Contents

Introduction.....	1355
I. Due Process Problems with Sampling	1359
A. The Statutory Framework.....	1359
B. Sampling: A Solution with Its Own Problems.....	1361
C. Statutory Analysis	1366
II. The FCA: Patchwork Enforcement, Deterrence Problems, and Misaligned Incentives	1367
III. Evaluating the Due Process Problem with Sampling	1371
A. Private Interests	1374
B. Risk of Error.....	1377
C. Interests of the Government or Relator	1382
1. Direct financial interests	1382
2. Signaling and deterrence interests.....	1383
3. Negative impacts on government and relator interests	1387
D. Balancing	1389
IV. Guidelines for Implementing Statistical Analysis	1390
A. Modified Bellwether Option	1390
B. Class Action Analog.....	1393
Conclusion.....	1396

Introduction

Few issues incite furor and partisanship in the present political climate more than federal health care spending. Indeed, in recent years Republican leaders have stated that Congress should place Medicare spending in its crosshairs, citing growing concerns about cost and debt.¹ While cutting the program is one way to reduce costs, fixing systemic difficulties in prosecuting Medicare fraud could also save the government billions without sacrificing coverage for senior citizens. According to the Centers for Medicare and Medicaid Services (CMS), fraudulent submissions to Medicare cost taxpayers billions of dollars per year.² In order to protect the public from this rampant abuse, as well as other fraudulent schemes, Congress has passed a suite of statutes designed to punish fraud, graft, and corruption.³ The False Claims Act (FCA) is one such statute.⁴ The FCA creates civil liability for any person who “knowingly presents, or causes to be presented, a false or fraudulent claim for payment or approval,” or “knowingly makes, uses, or causes to be made or used, a false record or statement material to a false or fraudulent claim.”⁵ This statutory provision is often used to file suits combating Medicare fraud,⁶ and represents a potential solution to the Medicare crisis.

-
1. See, e.g., Cameron Joseph, *McConnell Calls for Cutting Government Programs to Deal with “Disturbing” Debt*, BUS. INSIDER (Oct. 16, 2018, 5:11 PM), <https://perma.cc/GFM2-MSU2> (“Senate Majority Leader Mitch McConnell . . . called on Congress to rein in major government programs like Medicare, Medicaid and Social Security in order to slow America’s spiraling national debt on Tuesday, ignoring the fact the tax plan he recently passed has further grown that number.”); Jeff Stein, *Ryan Says Republicans to Target Welfare, Medicare, Medicaid Spending in 2018*, WASH. POST: WONKBLOG (Dec. 6, 2017), <https://perma.cc/M2S3-MAH5>.
 2. See MEDICARE LEARNING NETWORK, CTRS. FOR MEDICARE & MEDICAID SERVS., ICN MLN4649244, MEDICARE FRAUD & ABUSE: PREVENT, DETECT, REPORT 4 (2019), <https://perma.cc/E55C-KWDT>; *Health-Care Fraud: The \$272 Billion Swindle*, ECONOMIST (May 31, 2014), <https://perma.cc/4FZL-WA2P> (estimating that in 2012, “fraud (and the extra rules and inspections required to fight it) added as much as \$98 billion, or roughly 10%, to annual Medicare and Medicaid spending”).
 3. See MEDICARE LEARNING NETWORK, *supra* note 2, at 7-10 (describing the statutes designed to combat fraud).
 4. See *id.* at 7 (describing the FCA as protecting “the Federal Government from being overcharged or sold substandard goods or services”); see also False Claims Act, ch. 67, 12 Stat. 696 (1863) (codified as amended at 18 U.S.C. § 287 (2017); and 31 U.S.C. §§ 3729-3733 (2017)).
 5. 31 U.S.C. § 3729(a)(1)(A)-(B).
 6. See MEDICARE LEARNING NETWORK, *supra* note 2, at 7 (explaining that the FCA targets knowing submission of false claims for payment or reimbursement from the federal government). For example, in 2014 the Department of Justice (DOJ) recovered roughly \$5.7 billion under the FCA, with roughly \$2.3 billion coming from health care fraud cases. Lori L. Pines et al., *DOJ + FCA = Trouble for Corporations: False Claims Act Morphs into All-Purpose Anti-Fraud Tool*, CORP. COUNS. BUS. J. (Mar. 31, 2015), <https://perma.cc/GZJ4-TQK5>.

The specter of big data produces anxiety in the legal community, but fears about “trial by statistics” are exaggerated. Individualized adjudication is a bedrock of American jurisprudence, and the U.S. Supreme Court has repeatedly recognized this principle to be protected by the Due Process Clause.⁷ Statistical forecasting techniques use virtually the opposite approach to traditional legal analysis, focusing on patterns and relationships between a number of similar situations to construct inferences rather than making individualized findings.

Perhaps unsurprisingly, the Supreme Court has looked askance at plaintiffs’ efforts to introduce statistical evidence to prove liability in class action cases. For example, the Court rejected statistical evidence about Wal-Mart’s discriminatory labor practices in the landmark case *Wal-Mart Stores, Inc. v. Dukes*,⁸ seemingly foreclosing class action plaintiffs’ ability to submit statistical proof during the liability phase of a trial.⁹ But just five years later, the Court limited its holding in *Dukes*, holding in *Tyson Foods, Inc. v. Bouaphakeo* that class action plaintiffs may introduce statistical inference evidence when the evidence would prove an element of the claim in each individual plaintiff’s trial.¹⁰ While the Court relaxed its position on individualized adjudication to some extent, it carefully limited its holding in *Bouaphakeo* to just a subset of class action cases, leaving open the question whether plaintiffs could use statistical inference in any other context.¹¹

Demonstrating hospitals’, health care facilities’, or physicians’ liability for Medicare fraud under the FCA one payment at a time is unworkable where tens of thousands of Medicare payments are at issue, leading plaintiffs to attempt to introduce statistical evidence to prove liability.¹² So far, these attempts have met with mixed success, since the Supreme Court’s precedent remains fractured and plaintiffs do not file FCA claims as class actions.¹³

7. See, e.g., *Kerry v. Din*, 135 S. Ct. 2128, 2144 (2015) (holding, in the context of visa issuance decisions, that whenever individual rights or benefits are adjudicated, due process normally attaches to each individual decision).

8. See 564 U.S. 338, 356 (2011).

9. See Saby Ghoshray, *Hijacked by Statistics, Rescued by Wal-Mart v. Dukes: Probing Commonality and Due Process Concerns in Modern Class Action Litigation*, 44 LOY. U. CHI. L.J. 467, 499 (2012) (describing *Dukes* as supporting the idea that statistical extrapolations violate due process).

10. See 136 S. Ct. 1036, 1046 (2016).

11. See *id.*

12. See, e.g., *United States ex rel. Michaels v. Agape Senior Cmty., Inc.*, No. 0:12-3466-JFA, 2015 WL 3903675, at *8 (D.S.C. June 25, 2015), *aff’d in part, appeal dismissed in part*, 848 F.3d 330 (4th Cir. 2017); *United States ex rel. Martin v. Life Care Ctrs. of Am., Inc.*, 114 F. Supp. 3d 549, 570-71 (E.D. Tenn. 2014).

13. Compare *Martin*, 114 F. Supp. 3d at 556, 569-70 (accepting sampling to show liability with over 150,000 claims at issue), with *Michaels*, 2015 WL 3903675, at *1-2 (rejecting sampling to show liability with over 50,000 claims at issue).

Unfortunately, the lower courts have also rarely and inconsistently considered the application of statistical analysis to FCA liability, leaving open large questions about the viability of statistical techniques in such cases.

Unfortunately, as it stands now, the Department of Justice (DOJ) finds it exceedingly difficult to prosecute more than a handful of Medicare fraud cases when care providers habitually submit similar fraudulent claims. This leaves serial abusers free to squeeze the federal coffers, with little chance of punishment.¹⁴ While there is a *qui tam*¹⁵ provision allowing private plaintiffs to step in and prosecute Medicare fraud on the government's behalf,¹⁶ actually prosecuting such cases can cost more than a plaintiff would expect to recover. This is because an expert, usually a physician, must examine each purportedly fraudulent submission.¹⁷ Needless to say, this analysis is costly and may quickly swamp a *qui tam* plaintiff's expected gains,¹⁸ particularly in the common event that the defendant's alleged fraud was programmatic and continual.¹⁹ Nevertheless, randomly sampling from a large pool of potentially fraudulent claims to determine a hospital's or physician group's probability of liability presents a potential solution.²⁰ Given the Supreme Court's mixed messages in *Dukes* and *Bouaphakeo*, lower courts have little guidance in deciding whether to allow sampling.²¹ Unfortunately, sampling also presents a major problem

-
14. See S. REP. NO. 99-345, at 3 (1986) (“[M]ost fraud goes undetected due to the failure of Governmental agencies to effectively ensure accountability on the part of program recipients and Government contractors.”).
 15. A *qui tam* suit is defined as “[a]n action brought under a statute that allows a private person to sue for a penalty, part of which the government or some specified public institution will receive.” *Qui Tam Action*, BLACK’S LAW DICTIONARY (10th ed. 2014).
 16. See 31 U.S.C. § 3730(b)(1) (2017).
 17. See, e.g., *Michaels*, 2015 WL 3903675, at *1.
 18. See *id.* at *4 (explaining that the expected expense of expert review of all charts was between \$16 million and \$37 million, while the expected government recovery would only be \$25 million).
 19. See, e.g., *United States ex rel. Hutcheson v. Blackstone Med., Inc.*, 647 F.3d 377, 378-79 (1st Cir. 2011) (reversing a district court’s dismissal of a case involving a “nationwide” pattern of fraud); *United States ex rel. Spay v. CVS Caremark Corp.*, 913 F. Supp. 2d 125, 137-38 (E.D. Pa. 2012) (declining to dismiss a nationwide claim where the plaintiff alleged tens of thousands of counts of Medicare fraud); *United States ex rel. Thompson v. Columbia/HCA Healthcare Corp.*, 20 F. Supp. 2d 1017, 1031, 1049 (S.D. Tex. 1998) (denying summary judgment for the defendant on a claim of systematic Medicare fraud through inappropriate payment coding).
 20. Cf. *United States ex rel. Martin v. Life Care Ctrs. of Am., Inc.*, 114 F. Supp. 3d 549, 570-72 (E.D. Tenn. 2014) (approving of this solution).
 21. See Robert G. Bone, *Tyson Foods and the Future of Statistical Adjudication*, 95 N.C. L. REV. 607, 633 (2017) (discussing how lower courts have little guidance on how to implement *Bouaphakeo*).

under the Due Process Clause: Can a defendant be made to pay without a finder of fact determining *actual* liability for each count under the FCA?²²

Ultimately, these due process issues, discussed in Part I, present a clear set of questions that the courts should answer more definitively. First, should liability in FCA cases be proven for each claim, or does statistical sampling of a portion of claims suffice? Second, if sampling is allowed, under what conditions should it be allowed?

The rest of this Note addresses these questions, concluding that sampling is appropriate in cases against a single defendant with large numbers of claims at issue where the alleged fraud is systematic and the variability between claims is relatively limited. Part II unpacks the requirements for bringing a case under the FCA, as well as the application of due process precedent to FCA cases. Part III examines the characteristics of the cases where questions about sampling arise. Part IV discusses the procedural due process interests at stake in allowing statistical extrapolation. Part V discusses options for legal rules in light of this evaluation of the process due to defendants, examining the *Mathews v. Eldridge* balancing factors.²³

While commentators are increasingly grappling with these issues,²⁴ consideration of the actual process a statistical analyst would use to assess liability in FCA claims demonstrates that sampling need not offend due process. Perhaps unsurprisingly, more abstract considerations of “justice”²⁵ and high-level discussions of sampling²⁶ conclude that sampling violates due

22. See Ghoshray, *supra* note 9, at 499-502 (describing the liberty issues at stake in moving to sampling-based analysis of liability for class action litigation).

23. See 424 U.S. 319, 334-35 (1976).

24. See, e.g., Peter T. Thomas, Note, *Trial by Formula: The Use of Statistical Sampling and Extrapolation in Establishing Liability Under the False Claims Act*, 74 WASH. & LEE L. REV. ONLINE 103, 135 (2017) (arguing that the statutory text of the FCA bars sampling); Milene Vega, Note, *Should Statistical Sampling Be Used to Prove Liability Under the False Claims Act in Healthcare Fraud?*, 91 S. CAL. L. REV. 551, 589 (2018) (discussing the use of a modified bellwether approach to establish liability under the FCA); Christina Vlahos, Note, *When the Ends Do Not Justify the Means: The Application of Statistical Sampling to Determine Liability in False Claims Act Cases*, 90 ST. JOHN'S L. REV. 813, 837-38 (2016) (arguing that *Mathews* balancing precludes sampling in FCA cases). In contrast, the one analysis concluding courts should reconsider allowing extrapolation evaluates the issue without the benefit of the Supreme Court's guidance in *Bouaphakeo*. See Recent Case, *United States ex rel. Martin v. Life Care Centers of America, Inc., Nos. 1:08-cv-251, 1:12-cv-64, 2014 U.S. Dist. LEXIS 142660 (E.D. Tenn. Sept. 29, 2014)*, 128 HARV. L. REV. 2074, 2081 (2015). This analysis also largely focuses on the applicability of the reasoning in *Bigelow v. RKO Radio Pictures, Inc.*, 327 U.S. 251 (1946), to sampling. See Recent Case, *supra*, at 2078-81. It does not address the broader due process or statutory issues.

25. See, e.g., Vega, *supra* note 24, at 588 (“Part of the issue also centers on justice—it is not fair to determine knowledge as to thousands of claims from a small sampling.”).

26. See, e.g., Vlahos, *supra* note 24, at 840-43 (discussing general issues with statistical extrapolation).

process or does not comport with the scienter requirement in the FCA.²⁷ In contrast, I consider the question from the perspective of a specific forecasting strategy, which is how a well-informed court determining the constitutionality of extrapolation would approach the question. While this Note largely focuses on due process considerations, much of the analysis also responds to the claim that the FCA bars statistical analysis.²⁸

I. Due Process Problems with Sampling

Medicare fraud is usually prosecuted under the FCA because the statute makes unlawful the knowing submission of fraudulent or false claims to the government.²⁹ This Part unpacks the systemic problems hampering effective civil actions and deterrence under the statute, as well as the due process concerns with the existing solutions.

A. The Statutory Framework

A claim under the FCA proceeds in four parts: The plaintiff must identify specific claims made to the government, prove the claims or records are false or fraudulent, prove the defendant's knowledge of falsity, and, finally, prove the fraud or falsity to be material to the government's decision to pay the claims.³⁰ The Supreme Court recently clarified that the materiality requirement is meant to ensure that FCA claims go beyond "garden-variety breaches of contract or regulatory violations," and that violating minor or irrelevant regulations would not constitute material fraud.³¹ Traditionally, each false

27. See, e.g., *id.* at 844 ("In cases where statistical analysis would be used to determine defendants' liability, . . . defendants are denied the opportunity to defend themselves against each individual claim of fraud."); see also 31 U.S.C. § 3729(a)(1)(A)-(B) (2017).

28. Compare Joel D. Hesch & Mia Yugo, *Can Statistical Sampling Be Used to Prove Liability Under the FCA or Does Each Provision of the Statute Require Individual Proofs?*, 41 AM. J. TRIAL ADVOC. 335, 337-38 (2017) (suggesting the FCA does allow sampling), with Thomas, *supra* note 24, at 108 (suggesting it does not).

29. See 31 U.S.C. § 3729(a)(1)(A)-(B); Thomas Reilly, Comment, *The Extrapolation Conundrum: Finding a Unified Theory for the Use of Statistical Sampling in Medicare Fraud Cases Brought Under the False Claims Act*, 47 SETON HALL L. REV. 1103, 1104 (2017) ("With executive, administrative, and institutional efforts failing to prevent widespread health care fraud, the government increasingly relies on an old, though still very useful tool: the False Claims Act.").

30. See *United States ex rel. Martin v. Life Care Ctrs. of Am., Inc.*, 114 F. Supp. 3d 549, 565-70 (E.D. Tenn. 2014) (describing the elements of an FCA claim).

31. See *Universal Health Servs., Inc. v. United States ex rel. Escobar*, 136 S. Ct. 1989, 2003 (2016) (explaining the distinction between typical fraud claims and FCA claims).

claim allegation is proven separately, with specific documentary or expert evidence.³² Since the statute contains a scienter requirement, plaintiffs must prove actual or constructive knowledge.³³ Proving knowledge of fraud in Medicare reimbursement cases often requires expert testimony to show that a provider should not have ordered a procedure submitted for reimbursement, or did not perform the tests or procedures claimed.³⁴

The FCA contains two different enforcement mechanisms for pursuing those who knowingly submit fraudulent claims for government reimbursement or payment: First, the government can directly file a complaint against the tortfeasors;³⁵ second, there is a qui tam provision, allowing whistleblowers (termed “relators”) to file suit and collect a portion of the damages, usually between 15% and 30%, depending in part on whether the government intervenes in the suit.³⁶ Since qui tam plaintiffs “step into the shoes” of the government, the government can choose to intervene within sixty days to take control of a privately initiated claim.³⁷ In either case, victorious plaintiffs receive treble damages,³⁸ in line with Congress’s attempt to encourage suits and discourage fraud.

Some alleged FCA violations, particularly those in Medicare cases, involve patterns of fraudulent claims, often running into the tens of thousands.³⁹ For example, a hospital may routinely send patients with simple claims for unnecessary testing, creating thousands of unnecessary Medicare claims over the course of years or even decades.⁴⁰ Since Medicare cases require significant

32. See, e.g., *Martin*, 114 F. Supp. 3d at 565; see also *Vega*, *supra* note 24, at 558-60 (explaining the historical relationship between sampling and the FCA).

33. See 31 U.S.C. § 3729(a)(1)(A)-(B).

34. See, e.g., *United States ex rel. Michaels v. Agape Senior Cmty., Inc.*, No. 0:12-3466-JFA, 2015 WL 3903675, at *5 (D.S.C. June 25, 2015), *aff’d in part, appeal dismissed in part*, 848 F.3d 330 (4th Cir. 2017).

35. See 31 U.S.C. § 3730(a) (“The Attorney General diligently shall investigate a violation under section 3729. If the Attorney General finds that a person has violated or is violating section 3729, the Attorney General may bring a civil action under this section against the person.”).

36. See *id.* § 3730(b)(1) (“A person may bring a civil action for a violation of section 3729 for the person and for the United States Government. The action shall be brought in the name of the Government.”); *id.* § 3730(d)(1)-(2); see also 1 WEST’S FEDERAL ADMINISTRATIVE PRACTICE § 616 (West 2018) (“To aid in the battle against fraud, the Act’s ‘qui tam’ provisions allow a person (the relator) to bring a civil action on behalf of himself as well as the United States for the penalty recovered in court.”).

37. See 31 U.S.C. § 3730(b)(2).

38. See *id.* § 3729(a)(1).

39. See, e.g., *Michaels*, 2015 WL 3903675, at *1 (describing a fraud claim alleging between 53,000 and 62,000 counts under the FCA).

40. See, e.g., *United States v. Friedman*, No. 86-0610-MA, 1993 WL 13957433, at *1 (D. Mass. July 28, 1993) (“The United States alleges that one type of false claim was the hospitali-
footnote continued on next page”).

expert testimony in order to classify each claim as fraudulent, it may be the case that such large-scale patterns of fraud are simply too expensive to litigate⁴¹: The price of having experts individually review every file would be greater than even the treble damages the government could recover.⁴² This problem is compounded in qui tam suits where the government chooses not to intervene—the plaintiffs must bear the costs of the suit,⁴³ but they may receive insufficient compensation to carry it through to its conclusion.⁴⁴ This creates an unfortunate dynamic: the larger the scope of the fraud, the more significant the perverse incentive to settle or avoid bringing suit because of the costs of proving each count. The expected recovery only increases by roughly 25% of the increase in the full expected payoff, because qui tam plaintiffs only receive a quarter of the returns from successful cases. As a result, large-scale fraud cases are likely underdeterred,⁴⁵ contributing to ballooning budget outlays to Medicare year after year.

B. Sampling: A Solution with Its Own Problems

Fortunately, enterprising plaintiffs have already suggested a potential solution: statistical sampling to demonstrate liability. Sadly, this solution is mired in controversy.⁴⁶ In this context, sampling means that rather than prove liability or damages for each claim, the plaintiff selects a random and representative subset of all claims and only presents specific proof about that subset of claims; this proof is then extrapolated to the remaining claims.⁴⁷ Of course, each claim will not be identical, since statistical proof would hardly

zation and office treatment of patients where such services were not medically necessary.”), *vacated* (D. Mass. Sept. 24, 1993).

41. See *Michaels*, 2015 WL 3903675, at *1, *5 (describing a large-scale pattern of fraud that would cost more in litigation expert fees than the expected recovery).
42. See *id.* at *1 (estimating the cost to review each claim file at \$1,600 to \$3,600).
43. See 31 U.S.C. § 3730(d)(2) (granting costs to the qui tam plaintiff when the government does not intervene, presumably in recognition of the fact that the plaintiff must bear costs of bringing suit).
44. See, e.g., *Michaels*, 2015 WL 3903675, at *5 (estimating a projected return of \$25 million, but expected costs of up to \$37 million).
45. See *United States ex rel. Michaels v. Agape Senior Cmty., Inc.*, 848 F.3d 330, 334-35 (4th Cir. 2017) (discussing the issues large-scale fraud claims create); Pamela Bucy et al., *States, Statutes, and Fraud: A Study of Emerging State Efforts to Combat White Collar Crime*, 31 CARDOZO L. REV. 1523, 1539 (2010) (discussing the general view in the context of state FCA qui tam provisions that qui tam suits do not contribute to deterrence).
46. See *United States ex rel. Wall v. Vista Hospice Care, Inc.*, No. 3:07-cv-00604-M, 2016 WL 3449833, at *11 (N.D. Tex. June 20, 2016) (rejecting the use of sampling because of the potential unreliability of the evidence, which raised issues similar to those in *Dukes*).
47. See *United States ex rel. Martin v. Life Care Ctrs. of Am., Inc.*, 114 F. Supp. 3d 549, 566 (E.D. Tenn. 2014) (describing this process).

be necessary if they were.⁴⁸ Instead, identifiable characteristics of the cases are utilized to assess the relationship between claims in the randomly selected sample and claims outside the sample. For example, one might compare cases with similar injuries or Medicare claim codes, or claims that come from the same hospital. While courts have allowed statistical evidence of liability in both FCA and non-FCA cases,⁴⁹ in many others, courts have denied this same kind of evidence.⁵⁰

Courts ruling on the use of statistical sampling in determining liability confront significant due process concerns. Courts are primarily concerned about the fairness of using sampling techniques in the judgment process, and in particular, whether finders of fact can use sampling to render an accurate decision on each individual claim.⁵¹ There are three particular concerns: (1) that statistical inference techniques erroneously classify behavior as creating liability, where a case-by-case analysis would not;⁵² (2) that statistical

48. *See id.* at 566-67 (explaining that precisely identical observations are not required as long as the sample is representative).

49. *See, e.g.*, *Tyson Foods, Inc. v. Bouaphakeo*, 136 S. Ct. 1036, 1046 (2016) (accepting statistical evidence in a limited set of class actions); *United States v. Rogan*, 517 F.3d 449, 453 (7th Cir. 2008) (similar); *United States ex rel. Ruckh v. Genoa Healthcare, LLC*, No. 8:11-cv-1303-T-23TBM, 2015 WL 1926417, at *4 (M.D. Fla. Apr. 28, 2015) (observing no ban on expert testimony about statistical evidence in *qui tam* actions); *United States ex rel. Martin v. Life Care Ctrs. of Am., Inc.*, Nos. 1:08-cv-251 & 1:12-cv-64, 2014 WL 4816006, at *18-19 (E.D. Tenn. Sept. 29, 2014); *United States v. Fadul*, No. DKC 11-0385, 2013 WL 781614, at *14 (D. Md. Feb. 28, 2013).

50. *See, e.g.*, *Wal-Mart Stores, Inc. v. Dukes*, 564 U.S. 338, 356 (2011) (rejecting statistical evidence in a class action labor dispute); *United States ex rel. Crews v. NCS Healthcare, Inc.*, 460 F.3d 853, 856-57 (7th Cir. 2006) (rejecting rudimentary statistical evidence in support of a claim alleging a pharmacist's large-scale scheme to recycle drugs); *United States ex rel. El-Amin v. George Washington Univ.*, 533 F. Supp. 2d 12, 31 n.9 (D.D.C. 2008) (requiring plaintiffs alleging a pattern of anesthesiology fraud to provide evidence in support of each alleged false claim); *United States v. Friedman*, No. 86-0610-MA, 1993 WL 13957433, at *1, *3 n.2 (D. Mass. July 28, 1993) (declining to adopt statistical evidence offered by the plaintiff regarding an alleged pattern of Medicare fraud for inpatient services), *vacated* (D. Mass. Sept. 24, 1993).

51. *See In re Estate of Marcos Human Rights Litig.*, 910 F. Supp. 1460, 1467 (D. Haw. 1995) (describing the fairness concerns with allowing statistical extrapolation for proof of liability), *aff'd sub nom. Hilao v. Estate of Marcos*, 103 F.3d 767 (9th Cir. 1996).

52. *See id.* at 1467-68 (raising and rejecting the argument that case-by-case analysis leads to more accurate classification).

approximation cannot properly capture the FCA's scienter requirement;⁵³ and (3) that the sample will be unrepresentative of the whole population of cases.⁵⁴

While these concerns are valid, they underestimate the power of modern statistical inference techniques. Statistical sampling utilizes a random subset of claims to generate predictions about the overall population of claims; if there are 60,000 claims against one defendant, the plaintiff may sample as few as 300.⁵⁵ In order to show that the subset of claims is representative of all the claims, statisticians typically compare the means of the characteristics under examination—such as type of injury, cost of claims, or gender and race of patients—between the sample and the overall population; to show that the sample is generally representative of the population, the means of these important observable characteristics should be similar between the sample and the population.⁵⁶ This requires coding a database of the full universe of claims based on these characteristics, meaning the general characteristics of all claims a plaintiff alleges are false need to be coded. This coding would not require expert knowledge.⁵⁷ An expert would then conduct what is called a “forecasting analysis,”⁵⁸ in which she extrapolates from the observed relationship between relevant characteristics under the FCA, like severity of injury and Medicare payment code selected, within the sample cases to determine whether liability was likely in the remaining cases.⁵⁹ This analysis would predict, based on the observable characteristics described above, the

-
53. See *United States ex rel. Wall v. Vista Hospice Care, Inc.*, No. 3:07-cv-00604-M, 2016 WL 3449833, at *11-12 (N.D. Tex. June 20, 2016) (holding that individualized proof is required to establish scienter in Medicare FCA cases where physicians are alleged to have improperly diagnosed patients).
54. See *id.* at *13 (“[I]n order to fairly and reliably draw . . . an inference’ from a sample, ‘the sample must be randomly selected and . . . representative of the whole.’” (alterations in original) (quoting *United States v. Pena*, 532 F. App’x 517, 520-21 (5th Cir. 2013))).
55. See *United States ex rel. Martin v. Life Care Ctrs. of Am., Inc.*, 114 F. Supp. 3d 549, 556 (E.D. Tenn. 2014) (describing a case similar to this example).
56. See Elizabeth A. Stuart, *Matching Methods for Causal Inference: A Review and a Look Forward*, 25 STAT. SCI. 1, 1-3 (2010) (describing the proper construction of a comparison group so that it is representative of the underlying population).
57. To be sure, some details of the coding process might be the subject of some dispute. These disputes could be settled under the *Daubert* standard, as discussed in Part III below. For a basic description of the process involved, see Neil Liberman, *Decision Trees and Random Forests*, TOWARDS DATA SCI. (Jan. 26, 2017), <https://perma.cc/LHT9-ZXBX>.
58. Forecasting analysis is distinct from regression because it attempts to predict *outside* of the sample, whereas regression analysis minimizes prediction errors *within* the sample. In other words, forecasting analysis provides insight prospectively, while regression is largely retrospective. See Patricia B. Cerrito, *The Difference Between Predictive Modeling and Regression 1-2* (2008), <https://perma.cc/HHG6-FUBN>.
59. See Sujit Singh, *What Is Statistical Forecasting?: A Snowfall-Based Explanation*, ARKIEVA (Apr. 13, 2018), <https://perma.cc/7VP9-KC3H> (describing forecasting analysis).

percentage of claims in which a jury or court would find liability. In other words, a statistician could take a few hundred claims and project whether a jury would find liability in thousands of cases.

This forecasting analysis could take a number of forms, each with advantages and disadvantages. Since the goal is to sort between cases where liability for false claims would attach and cases where it is not likely to attach, the Random Forest method is one such useful statistical technique,⁶⁰ although many others would work.⁶¹ The Random Forest has a distinct advantage: The end product can be a table classifying results as “true positive, false positive, true negative, or false negative.” Since the due process concern is that defendants are forced to pay claims that are “false positives,” this output corresponds directly to the due process question of how often statistical inference erroneously suggests liability, because it suggests how frequently such false positives are likely to occur.⁶² In this method, individual decision trees regarding each element of an FCA claim would be repeatedly tested on portions of the sample of claims, replacing and redrawing different portions of the sample each time.⁶³ A decision tree allows the model to answer a series of sequential questions, which in the FCA context would be related to whether

60. See Tin Kam Ho, *Random Decision Forests*, 3 PROC. INT’L CONF. ON DOCUMENT ANALYSIS & RECOGNITION 278, 278 (1995) (“Decision-tree classifiers are attractive because of their many advantages—the idea is intuitively appealing, training is often straight-forward, and best of all, classification is extremely fast.”).

61. Random Forest is a particular type of classification model from a more general set of models called Classification and Regression Trees for Machine Learning (CART). See Houston H. Stokes, *Model Building Using Nonlinear Nonparametric Methods: Revised Chapter 17 in Specifying and Diagnostically Testing Econometric Models* (Edition 3), at 3-4 (Mar. 6, 2012) (unpublished manuscript), <https://perma.cc/AG8S-B5AN> (“A random forest model uses bagging to improve the performance of a CART type model.”). These models all tackle the problem of classification of discrete choice (rather than choices over continuous variables) using a “tree” structure to impose order on the problem of choice. See *id.* at 4. This is not the *only* way to conduct such an analysis; it is just one that fits with the structure of the FCA, since the statute has four elements that each must be established by a preponderance of the evidence. See, e.g., *United States ex rel. Martin v. Life Care Ctrs. of Am., Inc.*, 114 F. Supp. 3d 549, 565-70 (E.D. Tenn. 2014) (describing the elements of an FCA claim). Alternative specifications exist to relax this assumption, including the Generalized Additive Model, the Projection Pursuit model, and many other recursive covering analytical tools. See Stokes, *supra*, at 2-3, 24 (describing these various options in detail). These techniques are particularly useful in medical cases because they are quite accurate, although they can be difficult to interpret. See Deven R. Desai & Joshua A. Kroll, *Trust but Verify: A Guide to Algorithms and the Law*, 31 HARV. J.L. & TECH. 1, 53 (2017).

62. See Stokes, *supra* note 61, at 19-20 (describing how Random Forest models use a subset of data to predict which choice is most likely among a set of discrete alternatives).

63. See *id.*

the defendant is liable.⁶⁴ These decision trees would mirror the elements of the FCA, including decisions like “is this knowingly fraudulent conduct?” and “was this a material misrepresentation?” The process of repeated testing “trains” the model about the relationship between characteristics—like injury type and treatment—and decision outcomes,⁶⁵ which would be whether liability is likely under each component of the FCA.

This process is a specific application of a statistical technique called “bootstrapping,” where repeated estimates are used to reduce error in estimations.⁶⁶ In other words, repeated simulations are used because the sample statistics themselves are normally distributed, with the true population parameter as the mean.⁶⁷ Therefore, repeated estimates should reduce the noise in sample statistics, without biasing the estimates toward, or away from, suggesting liability. This process, while obviously exceedingly complicated theoretically, is well suited to the classification problem embedded in the FCA.⁶⁸ The end result of its proper execution would be a table, estimating for the complete population the numbers of cases where liability should attach, the number where it should not, and, most importantly, the rate of false positives and false negatives that can be expected.⁶⁹ While this is certainly a mathematically involved process, and by no means the sole model available, the important legal takeaway is this: Models can quantify expected false positives and negatives, giving courts a hook to consider whether sampling is an appropriate method to determine liability when massive numbers of claims are grouped together.⁷⁰ Experts in this area are typically professional economists with doctoral degrees who specialize in applied econometrics.

64. See Liberman, *supra* note 57 (“Trees answer sequential questions which send us down a certain route of the tree given the answer.”).

65. See Stokes, *supra* note 61, at 19.

66. See *id.*; see also Leo Breiman, *Random Forests*, 45 MACHINE LEARNING 5 (2001) (applying this bootstrapping concept to the analysis of categorical choices).

67. See generally Houston H. Stokes, *Regression Analysis with Appropriate Specification Tests: Revised Chapter 2 in Specifying and Diagnostically Testing Econometric Models* (Edition 3), at 2-2 to -76 (Dec. 16, 2012) (unpublished manuscript), <https://perma.cc/FQ3X-AVLW> (explaining the bootstrapping technique).

68. See Stokes, *supra* note 61, at 19.

69. See *id.* at 19-29 (explaining the process for implementing a Random Forest model in statistical packages).

70. See Desai & Kroll, *supra* note 61, at 53 (describing the Random Forest model as one of several models that are highly accurate in the context of medical decisionmaking); Stokes, *supra* note 61, at 3 (“The random forest technique is especially suitable for classification problems involving many possible outcomes.”).

C. Statutory Analysis

While this Note largely focuses on the constitutional due process challenges of sampling, there remains some doubt whether the FCA bars statistical sampling as a statutory matter.⁷¹ Proponents of this reading torture the statute's text and extricate it from its context, exhibiting the worst excesses of textualism. For example, Peter Thomas has confused singularity and particularity in the statute.⁷² He is correct that the statute creates liability for each individual claim,⁷³ but that does not inexorably require particular proof about knowledge of falsity in each particular submission. In fact, the Court made this clear in its decision in *Bouaphakeo*, in which it pointed out that an expert report about the average length of time it would take to put on and take off a uniform would be probative evidence in each particular worker's case that the jury could have and should have considered when determining liability for underpayment.⁷⁴ Whether this would suffice to sustain liability on its own depends on the precision of the liability forecast. If liability could be perfectly categorized, then it would provide perfect particular proof. In other words, my suggestion that judges pay close attention to variance in determining whether to allow sampling resolves the concern that this textualist reading of the statute raises.⁷⁵ The purpose of sampling, then, is not to *avoid* consideration of each individual claim, but to leverage claims' commonalities to avoid *duplicative* consideration of repetitive characteristics of the claims.⁷⁶

71. See Thomas, *supra* note 24, at 135-36 (arguing that the language of the FCA bars statistical sampling).

72. See *id.* at 135 ("The use of singular words here shows that the source of FCA liability is the individual claim.").

73. See *id.*

74. See 136 S. Ct. 1036, 1047-48 (2016).

75. It is worth noting that a number of judges read the statute more purposively and considerably more broadly than Thomas suggests. See, e.g., *United States v. Griswold*, 24 F. 361, 363, 366 (D. Or. 1885) ("It is intended to protect the treasury against the hungry and unscrupulous host that encompasses it on every side, and should be construed accordingly."). Thomas's analysis seems to take for granted that textualism is a superior modality of statutory interpretation, which is up for debate. Additionally, this issue was considered and rejected in *Martin*. See Lester J. Perling & Jamie B. Gelfman, *Statistical Sampling in the Medicare Program: How to Use It and How to Challenge It*, in *HEALTH LAW HANDBOOK* 639, 656 (Alice G. Gosfield ed., 2018 ed.) (describing how the court in *Martin* "found that the FCA did not specifically preclude the use of statistical sampling").

76. *But see* Thomas, *supra* note 24, at 138 (suggesting that statistical extrapolation circumvents individualized determinations of liability, in violation of the FCA's text).

II. The FCA: Patchwork Enforcement, Deterrence Problems, and Misaligned Incentives

Although sampling is commonly an issue in class action suits⁷⁷ and used frequently in estimates of damages,⁷⁸ it is a relatively recent addition to plaintiffs' arsenal in FCA claims.⁷⁹ This Part examines the handful of recent cases in which plaintiffs sought to introduce statistical evidence to prove liability where individual trials for each claim would be impracticable. Ultimately, the precedents are muddled and underdeveloped, suggesting a need for additional attention to this doctrine.

FCA plaintiffs typically seek to introduce evidence of similarities between claims when individual adjudication of every claim would be impracticable. For example, in *United States ex rel. Michaels v. Agape Senior Community, Inc.*, qui tam plaintiffs, or "relators,"⁸⁰ brought an action on behalf of the government against a senior center, Agape Senior Community, alleging a pattern of Medicare fraud.⁸¹ The scale of the fraud in *Michaels* was so large that there was dispute about the number of claims that *could* be alleged, although there was some agreement that the number of claims was between 53,000 and 62,000.⁸² In order to prove that Agape was defrauding Medicare, Michaels planned to present expert testimony from a physician, who would testify as to the knowingly inappropriate care Agape provided in order to earn higher reimbursements. The cost of reviewing each file was pegged at \$400 per hour and each file was estimated to require as many as nine hours of review; this estimate did not include the cost of court appearances.⁸³ The court denied plaintiffs' request to allow an expert to construct a random sample of claims in order to jointly try a large percentage of the claims without proffering

77. See Hillel J. Bavli & John Kenneth Felter, *The Admissibility of Sampling Evidence to Prove Individual Damages in Class Actions*, 59 B.C. L. REV. 655, 657 (2018) ("Putative class plaintiffs have proposed methods to prove classwide liability or damages by proffering sampling evidence, contending that common questions predominate over individual questions notwithstanding the existence of differences among class members' claims.").

78. See *id.* at 669-74 (discussing the historical development of damages estimates in class action cases).

79. See *United States ex rel. Martin v. Life Care Ctrs. of Am., Inc.*, 114 F. Supp. 3d 549, 560 (E.D. Tenn. 2014).

80. See 1 WEST'S FEDERAL ADMINISTRATIVE PRACTICE, *supra* note 36, § 616 ("To aid in the battle against fraud, the Act's 'qui tam' provisions allow a person (the relator) to bring a civil action on behalf of himself as well as the United States for the penalty recovered in court.").

81. See *United States ex rel. Michaels v. Agape Senior Cmty., Inc.*, No. 0:12-3466-JFA, 2015 WL 3903675, at *1 (D.S.C. June 25, 2015), *aff'd in part, appeal dismissed in part*, 848 F.3d 330 (4th Cir. 2017).

82. See *id.*

83. See *id.* at *5.

individual proof as to the appropriateness of care in each case.⁸⁴ The reasoning was based on two considerations: First, the court identified potential concerns that might arise from denying Agape the right to challenge each individual claim; second, Agape claimed it planned to submit case-by-case evidence to challenge the validity of the statistical evidence as applied to each claim, purportedly negating any benefit of introducing that evidence at all.⁸⁵ This illustrates the most restrictive version of the possible rule—completely rejecting statistical inference to prove liability in FCA cases.

In a similar case, *United States ex rel. Wall v. Vista Hospice Care, Inc.*, a plaintiff brought claims under the FCA against Vista Hospice Care, alleging abuse of the Medicare Hospice Benefit (MHB) affecting approximately 12,000 patients.⁸⁶ The plaintiff sought to sample about 300 claims in order to establish Vista’s pattern of fraud against Medicare, but the district court was not amenable to her claim: The court ruled that each determination of eligibility for the MHB was too individualized to allow sampling, citing a familiar criticism of sampling as “Trial by Formula.”⁸⁷ This partly reflects a consideration unique to qualification for the MHB, which is subjectively determined;⁸⁸ however, the court also relied on *Bouaphakeo*, quoting its holding that “[t]he permissibility of statistical sampling turns on ‘the degree to which the evidence is reliable in proving or disproving the elements of the relevant cause of action.’”⁸⁹ The court specifically pointed out that “[n]o circuit has resolved whether statistical sampling and extrapolation can be used to establish liability in an FCA case where falsity depends on individual physicians’ judgment regarding individual patients.”⁹⁰ Thus, the court left open the possibility of introducing evidence of aggregate patterns when probative as to individual behavior.⁹¹ In both *Michaels* and *Wall*, the difficulty and cost of adjudicating claims numbering in the tens of thousands failed to overcome the

84. *See id.* at *6-8.

85. *See id.*

86. *See* No. 3:07-cv-00604-M, 2016 WL 3449833, at *1, *5 (N.D. Tex. June 20, 2016).

87. *See id.* at *11 (“Where the nature of the claim requires an individualized determination, that determination cannot be replaced by ‘Trial by Formula.’” (quoting *Wal-Mart Stores, Inc. v. Dukes*, 564 U.S. 338, 367 (2011))).

88. *See id.* at *11, *13 (explaining that MHB determinations are based on practitioner judgments about the risk of death in the following six months, rather than specific clinical guidelines).

89. *Id.* at *11 (quoting *Tyson Foods, Inc. v. Bouaphakeo*, 136 S. Ct. 1036, 1046 (2016)).

90. *Id.* at *12.

91. *See id.* at *12-14.

risk of erroneously finding liability on a subset of the claims; despite both courts recognizing these burdens, neither was willing to allow sampling.⁹²

There is only one case to date approving of extrapolation from a statistical sample to demonstrate liability in the FCA context, at least through the summary judgment phase: *United States ex rel. Martin v. Life Care Centers of America, Inc.*⁹³ *Martin* represents a significant deviation from previous precedent, and has allowed extrapolation for purposes such as determining damages and calculating prejudgment writs of attachment.⁹⁴ The defendant, Life Care, operated 200 nursing homes, which received over \$4.2 billion from Medicare over the five-year period from 2006 to 2011.⁹⁵ The government's evidence of scienter was common to all claims, and included a pattern of employee complaints and corporate retaliation.⁹⁶ The court denied Life Care's motion for partial summary judgment, rejecting its due process arguments and reasoning that "[d]efendant will be afforded due process by having the opportunity to depose the Government's expert, challenge the qualifications of the Government's expert, retain its own expert, and to present all of this evidence at trial."⁹⁷ The issue has not been tested on appeal. This represents the most permissive ruling to date on the admissibility of statistical inferences to demonstrate liability under the FCA.

Martin is not alone: A few courts have allowed sampling to prove liability in contexts outside the FCA, including fraud.⁹⁸ For example, in *United States v. Fadul*, the court found that a subsample of Dr. Fadul's fraudulent submissions to Medicare justified summary judgment for the government as to falsity of all

92. See *id.*; *United States ex rel. Michaels v. Agape Senior Cmty., Inc.*, No. 0:12-3466-JFA, 2015 WL 3903675, at *8 (D.S.C. June 25, 2015), *aff'd in part, appeal dismissed in part*, 848 F.3d 330 (4th Cir. 2017).

93. 114 F. Supp. 3d 549 (E.D. Tenn. 2014).

94. See Robert T. Rhoad et al., Feature Comment, *Extrapolation in FCA Litigation: A Statistical Anomaly or a Tactic Here to Stay?*, GOV'T CONTRACTOR, Jan. 13, 2016, ¶ 9, at 1 ("Prior to the *Life Care* ruling, sampling has rarely been used in FCA cases, and it has never been used at trial—without the consent of the defendant—to prove liability."). Although the D.C. Circuit did allow sampling to prove liability in *United States v. Krizek*, it was only because the defendant stipulated to the use of extrapolation. See 111 F.3d 934, 940-41 (D.C. Cir. 1997). The court never had a chance to rule on its legality as a matter of due process. See *id.* at 941 ("We conclude, therefore, that the [defendants] are bound by their agreement at trial that liability would be based on the seven-patient sample with damages to be extrapolated later."). The D.C. Circuit's opinion characterizes the concession as "counsel for the [defendants] not only agree[ing] to, but proffer[ing], the idea of going to trial based on a representative sample." *Id.* at 940.

95. *Martin*, 114 F. Supp. 3d at 551.

96. See *id.* at 555.

97. *Id.* at 570; see also *id.* ("[T]he Court finds that the use of statistical sampling and extrapolation in this action does not violate Defendant's due process rights.")

98. See Bavli & Felter, *supra* note 77, at 669-73 (discussing the short history of such cases).

2,000 claims, despite the denial of summary judgment on the scienter component of the FCA cause of action.⁹⁹ This was largely due to the commonality between the claims, which eliminated the need for individualized determinations about the veracity of the submissions.¹⁰⁰ This comports logically with the “risk of error” component of the *Mathews v. Eldridge* balancing test¹⁰¹: When the claims are systemically related, rather than uncorrelated, models leveraging their statistical associations are more accurate in their predictions of liability.¹⁰² Standard errors increase as the variance of the underlying population (or sample) increases, meaning that a less variable population will, all else equal, allow more accurate estimation of statistical association.¹⁰³

It is also possible to take a different approach, analyzing the savings in terms of scarce court time and resources produced by allowing statistical extrapolation. *In re Estate of Marcos Human Rights Litigation*¹⁰⁴ is an exemplar for this analysis. Addressing procedural due process challenges, the court determined that sampling allowed the plaintiffs access to swift and reasonable damages, which outweighed the small risk that sampling would force the estate of Ferdinand Marcos, the former dictator of the Philippines, to pay for nonmeritorious claims.¹⁰⁵ The court explicitly referred to the extremely large burden a requirement of individualized proof for all claims would impose on the judicial system: “The judicial and administrative time and costs of holding bipolar trials would also have been virtually, if not absolutely, prohibitive.”¹⁰⁶

99. See No. DKC 11-0385, 2013 WL 781614, at *8-10, *14 (D. Md. Feb. 28, 2013) (granting partial summary judgment for the government on a common law claim for payment by mistake of fact, using a sample of 152 records).

100. See *id.* at *14 (explaining that the technique produces a reasonable approximation when the claims share certain characteristics).

101. See 424 U.S. 319, 335 (1976) (describing the risk of error component of the test as “the risk of an erroneous deprivation of such interest through the procedures used”).

102. Standard errors are partially a function of variation in the sample. When variance is lower, standard errors will also decrease.

103. See Bavli & Felner, *supra* note 77, at 680-82 (describing general standards for statistical evidence).

104. 910 F. Supp. 1460 (D. Haw. 1995), *aff’d sub nom.* Hilao v. Estate of Marcos, 103 F.3d 767 (9th Cir. 1996).

105. See *id.* at 1461-62, 1465-68. There is some concern that *In re Estate of Marcos* is no longer good law in light of the Supreme Court’s discussion of the case in *Dukes*. See *Wal-Mart Stores, Inc. v. Dukes*, 564 U.S. 338, 348, 356-57 (2011). But the Court suggested in *Bouphakeo* that statistical evidence can be introduced in some circumstances, thus limiting the holding of *Dukes*. See *Tyson Foods, Inc. v. Bouphakeo*, 136 S. Ct. 1036, 1046, 1048 (2016).

106. *In re Estate of Marcos*, 910 F. Supp. at 1468; see *id.* (“Clearly it cannot be questioned that a one-on-one trial is more burdensome for the Court than an aggregate trial. The costs involved in conducting bipolar trials with 9,541 plaintiffs in this case would
footnote continued on next page”)

Additionally, since Marcos's estate had a clear private interest in disbursing as much of his inheritance in accordance with Marcos's wishes as possible, an aggregate trial supported his estate's interests, leading to *lower* overall damage awards than the projected awards from trying each claim individually.¹⁰⁷ Finally, the estate retained its right to a jury trial,¹⁰⁸ and the procedures utilized protected the estate's interests no worse than general class action procedures, leading the court to comment that "the whole jurisprudence of class action treatment of numerous claims supports the conclusion that the [estate] has suffered no due process violation."¹⁰⁹ However, serious questions remain as to whether such a procedure is appropriate outside of the class action context, particularly where there is a single named plaintiff (either the government or a relator). Thus, the next Part aims to give courts guidance in balancing due process considerations in the FCA liability context.

III. Evaluating the Due Process Problem with Sampling

The fact that there *are* false positives in sampling techniques raises familiar due process concerns: Is a process that results in attributing liability to a defendant with some rate of error a violation of the Due Process Clause? Courts typically deal with procedural due process concerns using the landmark test announced in *Mathews v. Eldridge*.¹¹⁰ This three-part test emphasizes (1) the magnitude of harm to protected individuals; (2) the risk of erroneous deprivation of life, liberty, or property stemming from utilization of the challenged process; and (3) the benefit to the government in its use.¹¹¹ The test as originally envisioned does not neatly apply to situations where both the

substantially surpass the costs of an aggregate trial which lasted only about one and one-half weeks.").

107. *See id.* ("[I]t appears that had [one-on-one trials] been utilized, each claim would have brought a higher judgment . . .").

108. *See id.* at 1468-69.

109. *Id.* at 1468.

110. *See* 424 U.S. 319, 334-35 (1976). It is worth noting that the courts considering the particular context of FCA sampling and liability have largely dealt with the issue in a summary manner, without specifically referencing the *Mathews* balancing test. Nevertheless, *Martin's* analysis contains the general ingredients of *Mathews* balancing, suggesting this is what was lying behind the district courts' analysis. *See* United States *ex rel.* Martin v. Life Care Ctrs. of Am., Inc., 114 F. Supp. 3d 549, 570 (E.D. Tenn. 2014). There is some uncertainty about whether *Mathews* is the correct test, but courts considering the issue have not cited an alternative test. *See, e.g.,* United States *ex rel.* Wall v. Vista Hospice Care, Inc., No. 3:07-cv-00604-M, 2016 WL 3449833, at *13 (N.D. Tex. June 20, 2016); United States *ex rel.* Michaels v. Agape Senior Cmty., Inc., No. 0:12-3466-JFA, 2015 WL 3903675, at *6-8 (D.S.C. June 25, 2015), *aff'd in part, appeal dismissed in part*, 848 F.3d 330 (4th Cir. 2017); *Martin*, 114 F. Supp. 3d at 570.

111. *See Mathews*, 424 U.S. at 334-35.

plaintiff and defendant are private parties, such as a qui tam suit under the FCA; however, because the plaintiff is *assigned* the government's claim under the FCA, a qui tam suit still retains the characteristics of state action.¹¹² Thus, courts typically analyze the interests of the *government* in place of the interests of the qui tam plaintiff when considering due process challenges to FCA litigation procedures.¹¹³ As a result, the due process framework in FCA cases still draws on the traditional *Mathews* balancing categories: (1) harm to the alleged defrauder from being held liable under the FCA; (2) the risk of erroneously condemning the defendant's behavior as fraudulent or subjecting him to punishment for innocent claims; and (3) the benefit to the government, including the pass-through benefit the relator receives, from the procedure used for determining liability.¹¹⁴

At least one court has specifically applied this *Mathews* framework to a non-FCA case that used sampling and statistical inference to demonstrate liability across a large number of claims. The court in *In re Estate of Marcos* confronted this issue in the context of a pattern of human rights abuses.¹¹⁵ The court expressed concern with its conclusion about the magnitude of misclassification—how frequently sampling resulted in misclassifying legitimate exercises of state power as human rights abuses—as well as the actual expense statistical techniques saved.¹¹⁶ These reservations about reliability informed future challenges to the use of sampling in determining liability.¹¹⁷ Now, any serious attempt to use sampling needs to confront these twin due process considerations.¹¹⁸

Finally, it is important to note that this framework is intended to be flexible, so that judges making crucial, life-altering decisions are not handcuffed by statistical inference. The U.S. Supreme has Court emphasized

112. See *Eisenberg v. Mathews*, 420 F. Supp. 1274, 1279 (E.D. Pa. 1976) (analyzing an FCA claim by weighing the government's interest against those of the private defendants, similar to *Mathews*).

113. See, e.g., *Martin*, 114 F. Supp. 3d at 570; *Eisenberg*, 420 F. Supp. at 1279.

114. See *Martin*, 114 F. Supp. 3d at 570 (discussing the possibility of erroneous deprivation of the defendant's right to individual adjudication of claims); *Eisenberg*, 420 F. Supp. at 1279 (discussing the government's interest in efficient antifraud protections, the defendant's interest in an accurate judgment, and the risk that statistical analysis would erroneously deprive the defendant of his interests).

115. See *In re Estate of Marcos Human Rights Litig.*, 910 F. Supp. 1460, 1461-62, 1465-68 (D. Haw. 1995), *aff'd sub nom.* *Hilao v. Estate of Marcos*, 103 F.3d 767 (9th Cir. 1996); *supra* notes 104-09 and accompanying text.

116. See *id.* at 1468. *But see id.* (explaining that these concerns might be overblown).

117. See *Bavli & Felner*, *supra* note 77, at 661 (citing *Tyson Foods, Inc. v. Bouaphakeo*, 136 S. Ct. 1036, 1046 (2016)).

118. See *Vlahos*, *supra* note 24, at 837-38 (arguing that *Mathews* balancing precludes sampling in FCA claims due to reliability concerns).

this flexibility in due process cases, stating that “[d]ue process,’ unlike some legal rules, is not a technical conception with a fixed content unrelated to time, place and circumstances.”¹¹⁹ Thus, some of the arguments laid out below in favor of sampling in FCA cases may not be applicable in particular cases, especially where there is a high risk of error. For example, sampling is likely not appropriate in the case where doctors in a large physician group are encouraged to opportunistically “upcode”¹²⁰ procedures, which involves charging for services not rendered. In such cases, because opportunism, rather than a programmatic rule, generates the upcoding,¹²¹ this behavior may not be amenable to sampling. Thus, a court may find that sampling in order to prove liability in upcoding cases denies due process to a defendant. In contrast, if a defendant uses an algorithm to determine billing codes, which generates upcoded bills in a specific segment of claims, the relationship between the putatively fraudulent claims will be programmatic and thus amenable to statistical analysis. Indeed, the provider would actually be using precisely the sort of algorithm a statistician would use to conduct a Random Forest-type analysis. Therefore, even if the general rules outlined below are not appropriate in all cases, courts should weigh the due process ramifications carefully to find the contexts where the benefits of extrapolation outweigh the costs.¹²²

Most judges are not statisticians, but applying the *Mathews* balancing test to statistical sampling requires them to weigh important statistical and economic evidence. Given the general dearth of judicial and legal training in both statistics and economics, this Part provides streamlined guidance for courts considering the due process ramifications of extrapolating from a sample to prove liability in a whole population of claims.

In order to determine whether statistical extrapolation comports with the requirements of the Due Process Clause, courts must consider the *Mathews* balancing requirements.¹²³ This Part considers each of the *Mathews* balancing factors in turn, concluding that potential misclassifications can be contained

119. *Cafeteria & Rest. Workers Union, Local 473 v. McElroy*, 367 U.S. 886, 895 (1961) (quoting *Joint Anti-Fascist Refugee Comm. v. McGrath*, 341 U.S. 123, 162 (1951) (Frankfurter, J., concurring)).

120. *See United States ex rel. Bennett v. Medtronic, Inc.*, 747 F. Supp. 2d 745, 749-50 (S.D. Tex. 2010) (categorizing as upcoding the practice of relabeling “stand-alone surgical ablations” as “open-chest procedures to obtain favorable Medicare reimbursement rates”).

121. *But see id.* (describing how the alleged fraud *was* systematically related to the nature of the surgical procedure).

122. This need not be a strict cost-benefit analysis. This language is just used to reflect the general weighing principle embedded in the *Mathews* balancing test.

123. *See supra* Part II.

through effective implementation of statistical techniques like the Random Forest,¹²⁴ although there may be cases where wide variation in claims produces significant errors that outweigh the benefits of sampling. This Part proceeds in three Subparts: first, an analysis of doctors' and hospitals' interests; second, a discussion of the risk of error posed by sampling; and finally, an explication of the government's interests in using sampling. It concludes by balancing these factors, arguing that the appropriate result would allow sampling to prove liability where there are systematic patterns of fraud and large numbers of false claims alleged. Although different sampling techniques would be more or less reliable, the due process analysis in this Part concerns the general constitutionality of sampling to determine FCA liability.

A. Private Interests

The main private interest at stake in the decision to allow extrapolation of liability using sampling is the risk of overcompensation of plaintiffs for FCA violations, including awards for claims that are not actually fraudulent.¹²⁵ Forcing the defendant to pay for claims a jury would not have rewarded if tried independently would clearly constitute an erroneous deprivation of property without due process.¹²⁶ The magnitude of this harm is unclear, as property interests tend to be among the least strictly protected under the *Mathews* balancing test.¹²⁷ In fact, the Supreme Court has narrowed the scope of protected interests, clarifying that "a benefit is not a protected entitlement if government officials may grant or deny it in their discretion."¹²⁸ At least in comparison to life and liberty interests, property interests are more easily overridden by competing government efficiency and property interests.

124. See *supra* Part I.

125. See Jay Tidmarsh, *Resurrecting Trial by Statistics*, 99 MINN. L. REV. 1459, 1470 (2015).

126. *But see id.* at 1467 ("[T]rial by statistics can smooth out aberrational jury awards in individual cases.").

127. See Shankar Ramamurthy, *A Wolf in Sheeps Clothing: Chicago Public Schools' Disguised "For Cause" Termination and the Due Process Implications*, 7 DEPAUL J. FOR SOC. JUST. 235, 245 (2014) ("Not all property interests trigger constitutional due process protection. . . . In its seminal procedural due process ruling, . . . the Court explained that 'specific benefits' refers to property rights for which individuals have 'a legitimate claim of entitlement.'" (quoting *Bd. of Regents of State Colls. v. Roth*, 408 U.S. 564, 576-77 (1972))).

128. See *Town of Castle Rock v. Gonzales*, 545 U.S. 748, 756 (2005) (citing *Ky. Dep't of Corr. v. Thompson*, 490 U.S. 454, 462-63 (1989)). This may apply to rights that can be denied within judicial discretion, given that *Town of Castle Rock* considered the property interest involved in the grant of a restraining order. See *id.* at 750-51. Since the right to a trial is subject to judicial discretion, the private interest involved in the FCA context is potentially diminished.

When the government is the plaintiff, the incentive to pursue nonmeritorious claims is relatively weak; the government does not have a profit motive, but instead primarily works to maximize social utility.¹²⁹ Of course, effective deterrence and damages are linked, but maximizing damages with no regard for justice would not comport with the government's enforcement mission. Additionally, the incentive to bring weak cases is dampened because government officials do not directly receive benefits based on the quantum of damages collected. Qui tam plaintiffs, however, may be motivated by profit to bring as many colorable claims as possible. When private plaintiffs internalize the costs of bringing each claim, this counteracts their profit motive, pushing relators to mirror the government's incentive to bring only strong claims to trial. Additionally, since qui tam plaintiffs only receive a portion of the recovery, their motivation to bring claims is already smaller than the government's.

Allowing sampling seemingly reverses this dynamic, eliminating the need to pay an expert for each claim. A truly random sampling technique would partially resolve this conundrum, providing an incentive for the relator to avoid diluting her pool of claims for fear that the estimated probability of liability will end up being too low for meritorious claims.¹³⁰ Additionally, it is unclear why it would be unreasonable to compensate claims with probability greater than zero of victory at trial, even if that probability were relatively low: The compensation for those claims would be in recognition of the risk that a reasonable jury could find for the plaintiff. Put another way, a judge could not properly decide the case as a matter of law, because perhaps 10% of juries would find for the plaintiff.¹³¹ This dynamic would produce a number of settlements. Given that the summary judgment standard is already designed to recognize that the draw of jurors is a stochastic process, and thus so too is the ultimate verdict, it is unclear why courts should adopt the pretense that jury trials are not probabilistic when it comes to sampling extrapolation.¹³²

129. See Bruno S. Frey & Alois Stutzer, Conference Paper, *Should National Happiness Be Maximized?* 5 (Univ. of Zurich Inst. for Empirical Research in Econ. Working Paper Series, Working Paper No. 306, 2007), <https://perma.cc/B7JP-AA7V>.

130. It may be possible to intentionally manipulate the set of claims before filing, but the defense could submit its own expert testimony demonstrating that the plaintiffs had intentionally manipulated the pool of claims. Although this could create some concerns about judges sorting out dueling expert testimony, this is hardly a novel concern in this area of the law.

131. See *Celotex Corp. v. Catrett*, 477 U.S. 317, 322-23 (1986) (laying out the modern standard for judgment as a matter of law on issues of fact).

132. The standard for summary judgment in civil matters, which implicitly builds in the concept of the reasonable juror, suggests that the probability of jurors rendering a particular verdict is the controlling question. See FED. R. CIV. P. 56(a); see also *id.* r. 50(a)(1).

Alternatively, another framing of the due process right at issue in private tort cases is the right to bring suit, which is not at issue in FCA cases.¹³³ Bound to this is usually a valorization of the participatory aspects of the law: Individual suits are valuable because they encourage individual engagement in the process.¹³⁴ As the court pointed out in *Martin*, defendants are still afforded the right to present their cases in court, including the opportunity to negate any expert testimony on extrapolation or even to provide their own statistical evidence.¹³⁵ Although there may be some concern about the ability of courts to interpret expert evidence, or of lay jurors to recognize it, the reality is that the statistical tools involved are already deeply embedded in our civil litigation system.¹³⁶ Even given these concerns, a court may maximize the accuracy of adjudication over many claims using extrapolation from a sample of claims.¹³⁷ Thus, the claim that defendants are denied the right to present their case seems tenuous at best.

A final justification for individually litigating each claim is information revelation, or “shining a light” on corruption or innocence. But this motivation is not necessarily sacrificed when sampling is utilized. While it is true that the plaintiff will not air most salient details of individual acts of corruption when utilizing sampling to prove liability, the spotlight should more likely be placed on the *pattern* of corruption; indeed, statistical sampling places the emphasis precisely on a pattern of conduct.¹³⁸ Thus, the defendant, in demonstrating that the general pattern of corruption does not exist, would effectively clear her good name, preserving the information function of trials.

133. See Martin H. Redish & Julie M. Karaba, *One Size Doesn't Fit All: Multidistrict Litigation, Due Process, and the Dangers of Procedural Collectivism*, 95 B.U. L. REV. 109, 132 (2015) (“Each claimant in a[] [multidistrict litigation] has an individually held, constitutionally protected property right at stake. . . . The ‘property’ at stake . . . [is] the right to sue to enforce a legally protected claim, even the unlitigated right to sue.”).

134. See *id.* at 135 (“[I]ndividual participation is inherently valuable . . . because it legitimizes the adjudicating entities in the minds of the litigants.”).

135. *United States ex rel. Martin v. Life Care Ctrs. of Am., Inc.*, 114 F. Supp. 3d 549, 570 (E.D. Tenn. 2014).

136. See Bone, *supra* note 21, at 657-61 (describing the various ways sampling fits into the extant civil litigation system).

137. See Hillel J. Bavli, *Sampling and Reliability in Class Action Litigation*, 2016 CARDOZO L. REV. DE NOVO 207, 213 (“In particular, I show that, for a class of N homogeneous claims, a court maximizes accuracy by sampling \sqrt{N} claims, rather than all N claims, for individual adjudication.”).

138. One of the main justifications for the FCA is bringing to light patterns of false claims against the government in order to root out the systematic draining of government coffers. See *United States v. Griswold*, 24 F. 361, 366 (D. Or. 1885) (“[The FCA] is intended to protect the treasury against the hungry and unscrupulous host that encompasses it on every side, and should be construed accordingly.”).

Overall, while there are several compelling private interests at stake in statistical extrapolation, they are largely dependent on showing that sampling will result in significant error. If defendants are simply deprived of Medicare fees which were based on false claims, it is hardly a significant concern. Thus, much of the analysis turns on the risk-of-error component of *Mathews* balancing.

B. Risk of Error

The primary risk of error in statistical sampling concerns potential misclassification of claims, particularly false positives—where liability is found that would not have been found in the absence of extrapolation.¹³⁹ While this concern is warranted, there is little to distinguish the application of extrapolation in FCA cases from the class action and mass tort contexts, where similar techniques are already utilized.¹⁴⁰ Additionally, there is little evidence that juries properly assign liability to cases in the first place,¹⁴¹ making the counterfactual baseline suspect as a marker of correct classification. Finally, most concerns about statistical analysis are overblown, particularly where the population is well defined, relevant characteristics are known and observable, and underlying variation is relatively small.¹⁴²

Additionally, some scholars argue that statistical analyses cannot properly capture the texture of a claim in the way that individually litigating in open court can; something ineffable about the phenomenon of an individual's day in court holds sway in the imagination of many commentators.¹⁴³ There is no evidence, however, that the ultimate result of diverging from this Platonic ideal of a trial actually prejudices the defendant. In fact, courts have suggested the opposite may occur: Aggregate trials can potentially *decrease* the amount of expected damages a defendant will pay.¹⁴⁴

139. See *Mathews v. Eldridge*, 424 U.S. 319, 335, 343-44 (1976) (describing the risk-of-erroneous-deprivation prong of the *Mathews* test).

140. See *Tyson Foods, Inc. v. Bouphekeo*, 136 S. Ct. 1036, 1041, 1048 (2016) (holding that statistical extrapolation was acceptable in a serial underpayment class action).

141. See Bruce D. Spencer, *Estimating the Accuracy of Jury Verdicts*, 4 J. EMPIRICAL LEGAL STUD. 305, 307 (2007) (“[T]he agreement rates [between jury and judge verdicts] for criminal cases, excluding hung jury cases, are . . . just under 80 percent. Such an agreement rate is not cause for complacency, given that agreement by chance alone would exceed 60 percent for each study.”).

142. See Desai & Kroll, *supra* note 61, at 52-53 (describing the accuracy of these methods).

143. See Vlahos, *supra* note 24, at 843-44.

144. See, e.g., *In re Estate of Marcos Human Rights Litig.*, 910 F. Supp. 1460, 1468 (D. Haw. 1995) (“There is no proof that the [estate] would pay less had damages been determined on a bipolar basis. In fact, as stated above it appears that had such a procedure been utilized, each claim would have brought a higher judgment . . .”), *aff'd sub nom.* *Hilao v. Estate of Marcos*, 103 F.3d 767 (9th Cir. 1996).

Nor is medicine “too subjective” for care decisions to be placed within a categorical, statistical framework. While care determinations do rely on subjective analyses in *individual* determinations, the *pattern* of care will exhibit repeated characteristics, which provide significant information about a physician’s or hospital’s views.¹⁴⁵ For example, a doctor must decide if a patient has six months or less to live in order for that patient to qualify for the MHB.¹⁴⁶ This is largely subjective in each individual case, although there are rough clinical guidelines.¹⁴⁷ Nevertheless, if a provider consistently approves the MHB for patients with certain ailments or characteristics that would not ordinarily be considered terminal cases, coupled with a persistent pattern of hospice stays greatly exceeding six months, that would be evidence of a pattern of fraud. While it is tempting to presume that the jury will weigh what was going on inside the doctor’s head to establish scienter, this is impossible. The jury will not know what the doctor was thinking, and thus will rely on evidence that *suggests* what the doctor’s state of mind was when making subjective determinations. Applying a quantitative methodology to that process only serves to cabin potential errors in judgment within a consistent framework.

Since large-scale FCA actions tend to name only a single defendant,¹⁴⁸ it is unclear that defendants have a significant interest in minimizing the risk of error in any particular judgment.¹⁴⁹ Instead, a defendant’s interest is in minimizing the aggregate judgment for all claims in a complaint. Extrapolation allows more accurate classification in the aggregate, even if the method will make larger mistakes with respect to any individual claim compared to trying each count individually before a jury.¹⁵⁰ This reduces the *relevant* risk of error, even if it increases the risk of *some* measurement of error; no part of the *Mathews* test suggests which measurement of error due process is to minimize.

145. See Tidmarsh, *supra* note 125, at 1467-68 (explaining that accuracy may be improved using statistical extrapolation, depending on whether the reference point is accuracy of the average or of each individual count).

146. See United States *ex rel.* Wall v. Vista Hospice Care, Inc., No. 3:07-cv-00604-M, 2016 WL 3449833, at *3-5 (N.D. Tex. June 20, 2016).

147. *But see* Vega, *supra* note 24, at 586-88 (arguing that medicine is too “subjective” to be amenable to statistical analysis).

148. See, e.g., Wall, 2016 WL 3449833, at *1; United States *ex rel.* Michaels v. Agape Senior Cmty., Inc., No. 0:12-3466-JFA, 2015 WL 3903675, at *1 (D.S.C. June 25, 2015), *aff’d in part, appeal dismissed in part*, 848 F.3d 330 (4th Cir. 2017); United States *ex rel.* Martin v. Life Care Ctrs. of Am., Inc., 114 F. Supp. 3d 549, 551 (E.D. Tenn. 2014).

149. See Tidmarsh, *supra* note 125, at 1467, 1484-85 (explaining that using statistical extrapolation may actually improve accuracy for the average claim).

150. See *id.* at 1468 (“Even if it does not increase the accuracy of individual awards, however, trial by statistics may do a better job of determining the aggregate liability of the defendant—a fact that again loops back to . . . deterrence.”).

Logically, minimizing average error is at least as good, especially given the private interest at risk of deprivation is overall claim cost.¹⁵¹

The idea that any plaintiff will be properly compensated is already a legal fiction—maintaining the illusion that compensation must remain in lockstep with some idealized count of damages is appealing, but out of touch with reality.¹⁵² Many meritorious claims will not be filed simply as a result of undercompensation, creating underdeterrence.¹⁵³ Many dubious claims will prevail on the strength of effective counsel and weak opposition. Decrying as error any deviation from an idealized counterfactual scenario in which each individual count is tried separately before a jury seems questionable, given that the more likely alternative to statistical extrapolation is often to not try a large portion of claims at all.¹⁵⁴

Sampling in order to establish damages or as part of the evidence in a given case is a fact of life for many tort claims.¹⁵⁵ Statistical estimates are necessarily imprecise; only having the underlying population parameter, which is not a value observed in any realistic factual scenario, would avoid this fuzziness in damages calculations.¹⁵⁶ As a result, the concept of avoiding overcompensation rests on a picture of the justice system that does not reflect the increasingly systematized reality of most government processes from which FCA claims spring. For the most part, these are not small shops submitting tens or hundreds of Medicare claims; they are sophisticated operations adopting

151. See *id.* at 1467 (“[T]rial by statistics can enhance accuracy—a positive feature both for those who believe that legal process should be as efficient as possible and for those who believe that procedure’s role is to enforce substantive rights as perfectly as possible.”).

152. See David A. Hyman & Charles Silver, *Medical Malpractice Litigation and Tort Reform: It’s the Incentives, Stupid*, 59 VAND. L. REV. 1085, 1104 (2006) (“[U]nder-compensation is the norm in the tort system, although victims with small claims are sometimes modestly overpaid.”).

153. See Scott DeVito & Andrew W. Jurs, “Doubling-Down” for Defendants: The Pernicious Effects of Tort Reform, 118 PENN ST. L. REV. 543, 593-94 (2014) (explaining this dynamic in the context of medical malpractice).

154. See *infra* Part III.C.

155. See, e.g., *United States v. Jones*, 641 F.3d 706, 712 (6th Cir. 2011) (“A statistical estimate may provide a sufficient basis for calculating the amount of loss caused by a defendant”); *United States v. Rogan*, 517 F.3d 449, 453 (7th Cir. 2008) (“Statistical analysis should suffice.”); *In re Chevron U.S.A., Inc.*, 109 F.3d 1016, 1019 (5th Cir. 1997); *United States v. Fadul*, No. DKC 11-0385, 2013 WL 781614, at *14 (D. Md. Feb. 28, 2013); *In re Estate of Marcos Human Rights Litig.*, 910 F. Supp. 1460, 1467-68 (D. Haw. 1995), *aff’d sub nom. Hilao v. Estate of Marcos*, 103 F.3d 767 (9th Cir. 1996); *Cimino v. Raymark Indus., Inc.*, 751 F. Supp. 649, 653 (E.D. Tex. 1990), *rev’d in part, vacated in part*, 151 F.3d 297 (5th Cir. 1998).

156. See James Dean Brown, *Sample Size and Statistical Precision*, SHIKEN (JALT Testing & Evaluation SIG, Tokyo, Japan), Aug. 2007, at 21, 21-22, <https://perma.cc/A9D8-NFA4> (explaining that samples are necessarily imprecise, whereas population parameters are exact).

systematic billing practices, submitting tens of thousands of claims.¹⁵⁷ Adopting the same systematic approach to punishing false claims reflects the reality of the pattern of behavior the FCA attempts to confront and deter. Even if the size of the institution is relatively small, techniques like the Random Forest approach make use of *repeated* estimation, reducing the need for large samples, and instead leveraging thousands of simulations of the same estimation with random draws of cases to determine liability.

Sampling also accurately reflects the risk-management tools most hospitals and nursing facilities already utilize in making their care determinations.¹⁵⁸ The increasing complexity and computerization of hospital data management mimics the data collection processes an effective statistical analysis would utilize. As a result, even if the risk of error from statistical analysis is large in some contexts, misclassification should be virtually nonexistent in the context of Medicare claims, especially for the large-scale frauds at issue in *Martin, Michaels, and Wall*.¹⁵⁹ In fact, such analyses may be the most effective way to replicate the decisionmaking process risk managers and care providers use in constructing their business models.

Finally, safeguards can be built into the system, ensuring that the already low risk of error diminishes further. For example, the defendant may request a *Daubert* hearing on any statistical extrapolation testimony, prior to its inclusion in the trial.¹⁶⁰ *Daubert* hearings are already standard practice for screening out potentially prejudicial or unreliable expert testimony.¹⁶¹ It is unclear why proof of falsity requires a higher standard of reliability than showing fault in a malpractice claim or demonstrating antitrust causation, both contexts in which statistical expert evidence is allowed.¹⁶² Given the existing academic institutions built around applied statistics and econometrics, there are certainly many experts available to testify regarding the validity and

157. See generally Gibson Dunn, 2018 Mid-Year False Claims Act Update 4-6 (2018), <https://perma.cc/WH6Y-SQ8T> (discussing general DOJ health care settlement trends).

158. See United States *ex rel.* Martin v. Life Care Ctrs. of Am., Inc., 114 F. Supp. 3d 549, 551 (E.D. Tenn. 2014) (describing the defendant's practice of making care decisions in part based on risk-management software).

159. See *supra* Part II.

160. See *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579, 589, 592-93 (1993) (holding that parties have the right to a hearing on the reliability and relevance of expert testimony prior to trial).

161. See *id.* at 595, 597; see also FED. R. EVID. 702 (codifying the *Daubert* standard).

162. See Rhoad et al., *supra* note 94, at 2-3 (describing the higher standard of proof required for FCA claims).

precision of sampling and forecasting techniques. If, however, additional safeguards are necessary, more tailored pretrial screening provides an effective avenue to secure them.¹⁶³

Even if courts believe the risk of error is generally large, courts should allow extrapolation when variance of classification¹⁶⁴ is low. When sample variance decreases, the standard error associated with statistical inferences generally declines, all else held equal.¹⁶⁵ Thus, even if courts are concerned about accuracy of inferences in general, confidence should be greater when variance is small.¹⁶⁶ Additionally, courts should look beyond the variance of the sample itself to the variance of the forecasted rates of classification into liability and nonliability groups; even if the characteristics of the claims have high variance, the classification analysis may not, because it leverages the additional power of repeated simulations.¹⁶⁷ While residual uncertainty about the estimate of variance itself will remain, a number of statistical tests allow sufficiently accurate estimates to confidently assess the risk of error in any given case.¹⁶⁸ Courts should make this determination within the *Daubert* framework to allow the defendant the ability to challenge a potentially prejudicial application of the standard before submitting complicated evidence

163. For cases in which courts have screened out potentially unreliable evidence before trial, see, for example, *United States ex rel. Ruckh v. Genoa Healthcare, LLC*, No. 8:11-cv-1303-T-23TBM, 2015 WL 1926417, at *3-4 (M.D. Fla. Apr. 28, 2015) (ruling in response to a motion in limine that there is no universal ban on statistical extrapolation in qui tam suits); *United States ex rel. Loughren v. UnumProvident Corp.*, 604 F. Supp. 2d 259, 260, 263, 269 (D. Mass. 2009) (excluding statistical evidence from an FCA case as unreliable after holding a *Daubert* hearing).

164. This refers to the output of a Random Forest-type stochastic classification model, which separates cases into categories of liability and nonliability, and flags false positives and false negatives.

165. See WILLIAM H. GREENE, *ECONOMETRIC ANALYSIS* 62 (8th ed. 2017) (explaining that standard error in a traditional ordinary least squares model is a function of sample size and sample variance).

166. See *id.* at 75 (explaining statistical confidence construction as a function of sample variability).

167. It is important to remember, as discussed in Part I.B above, that there are two distributions in question here: the distribution of FCA-relevant characteristics, like type of injury and sex of the patient, and the distribution of computed sample statistics, based on a regression model or a decision tree model like the Random Forest. It may be the case that the distribution of characteristics has a high variance, but repeated simulations significantly reduce the variance of computed relationships between the sample and likelihood of liability.

168. See *Comcast Corp. v. Behrend*, 569 U.S. 27, 35 (2013) (“Calculations need not be exact, but at the class-certification stage (as at trial), any model supporting a ‘plaintiff’s damages case must be consistent with its liability case’” (citation omitted) (quoting ABA SECTION OF ANTITRUST LAW, *PROVING ANTITRUST DAMAGES: LEGAL AND ECONOMIC ISSUES* 62 (2d ed. 2010))); see also GREENE, *supra* note 165, at 640-93 (describing simulation-based approaches to validating sample statistics and standard errors).

to the jury. At this stage, both plaintiffs and defendants could submit expert reports using a subsample of the claims to demonstrate whether the variance between claims is too large. The judge could then determine whether sampling would be sufficiently reliable for this set of claims.

C. Interests of the Government or Relator

This Subpart examines the interests of the government, or a relator on behalf of the government, in utilizing sampling to prove liability. There are two major parts to this analysis: direct financial interests in deterring fraud and indirect deterrence benefits. Each is considered in turn. This Subpart concludes by examining potential harms to government interests from allowing sampling.

1. Direct financial interests

The principal aim of the FCA is to deter fraud against the government; unsurprisingly, the primary interest of both the government and relators with respect to *Mathews* balancing is to prevent fraudulent claims or to recover damages from those claims as civil punishment.¹⁶⁹ Relators also have a statutory interest in the recovery of damages, unrelated to the punishment of fraud and return of taxpayer money.¹⁷⁰

The government has a limited budget to prosecute fraud, meaning that the efficient use of its resources is a critical interest. Inefficient deployment of funds decreases both the number of cases in which the government can intervene to assist relators and the number of independent investigations it can afford to conduct.¹⁷¹ Allowing extrapolation would give officials targeting fraud a mechanism to save scarce resources.

In addition to the direct benefits to plaintiffs, allowing sampling would dramatically reduce court costs associated with trying each false claim individually.¹⁷² Instead of requiring individual proof and review of tens of

169. See 131 CONG. REC. 8778 (1985) (statement of Rep. Stark) (describing the intent of the FCA as preventing piracy against the government); Vlahos, *supra* note 24, at 816-17 (describing the impetus for the FCA as rampant contracting fraud during the Civil War).

170. See 31 U.S.C. § 3730(d) (2017).

171. See Eric Topor, *Intervention in False Claims Act Lawsuits: Is It Make or Break?*, BLOOMBERG BNA (Apr. 24, 2017), <https://perma.cc/C4VM-4FBF> (“[T]he DOJ intervenes in only about 25 percent of whistle-blower cases alleging Medicare and Medicaid billing fraud . . .”).

172. See Rhoad et al., *supra* note 94, at 1-3 (describing how sampling has been used in cases involving so many claims that litigating on an individual basis would be impractical).

thousands of claims, as the court required in *Michaels*,¹⁷³ trials could be conducted in as little as “one and one-half weeks.”¹⁷⁴ While courts would need to consider a difficult *Daubert* motion related to the particular sampling method, the cost in time and effort would pale in comparison to the consideration of the reliability of expert reports for each individual claim. Court time is an underappreciated source of inefficiency; a full docket creates the incentive for judges to lean in favor of granting motions to dismiss or summary judgment, and compresses judicial decision time allocated to each count.¹⁷⁵ Thus, the effects of forcing burdensome individual trials create significant externalities for *all* plaintiffs; aggregation reduces both of these costs.

2. Signaling and deterrence interests

Underdetering fraud undermines the key purpose of the FCA: eliminating fraud. Unfortunately, the current state of the law is likely underdetering fraudsters. First, it is often too expensive to bring a claim, given the hefty costs for expert testimony and analysis, particularly in Medicare cases.¹⁷⁶ Allowing sampling would reduce these costs, proportionate to the required sample size.¹⁷⁷

The current expense of prosecuting FCA claims could create a perverse incentive for fraudsters to widen the scope of their fraud in order to make it too expensive to prove each individual count. Undertaking the due diligence required to file a complaint may require significant investments in order to avoid the specter of sanctions under Federal Rule of Civil Procedure 11,¹⁷⁸ not to mention the probability of losing one’s employment for filing a *qui tam* action against an employer.¹⁷⁹ Few will willingly sustain such costs without

173. See *supra* text accompanying notes 80-85.

174. See, e.g., *In re Estate of Marcos Human Rights Litig.*, 910 F. Supp. 1460, 1468 (D. Haw. 1995), *aff’d sub nom.* *Hilao v. Estate of Marcos*, 103 F.3d 767 (9th Cir. 1996).

175. See J. Alexander Tanford, *A Political-Choice Approach to Limiting Prejudicial Evidence*, 64 IND. L.J. 831, 852 (1989) (“Inefficiency is therefore usually included among the legitimate meanings of prejudice.”).

176. See, e.g., *United States ex rel. Michaels v. Agape Senior Cmty., Inc.*, No. 0:12-3466-JFA, 2015 WL 3903675, at *5 (D.S.C. June 25, 2015), *aff’d in part, appeal dismissed in part*, 848 F.3d 330 (4th Cir. 2017).

177. See Tidmarsh, *supra* note 125, at 1459-60 (“Trial by statistics allows the judge to hear just a fraction of the total number of cases, making the resolution of mass disputes with varying amounts of individual damages a realistic possibility.”).

178. See, e.g., *Pentagen Techs. Int’l Ltd. v. United States*, 172 F. Supp. 2d 464, 471-73 (S.D.N.Y. 2001) (granting sanctions for abuse of the FCA where plaintiffs had filed multiple frivolous *qui tam* suits); see also FED. R. CIV. P. 11.

179. Technically, the FCA contains a nonretaliation provision. See 31 U.S.C. § 3730(h) (2017). The history of such provisions is mixed due to the difficulty of enforcement. See, e.g.,

footnote continued on next page

some certainty of recoupment, although it is possible for repeat players to defray costs of any individual loss over many cases. Bellwether trials may counteract this incentive to the extent that they create settlement pressure, but defendants would be able to continue using the threat of litigating each individual claim even if plaintiffs successfully tried the first hundred or so claims out of many thousands.¹⁸⁰ Allowing extrapolation would resolve the problem of large patterns of fraud decreasing the incentive to file suit. Establishing credible statistical inferences requires a smaller percentage of cases with larger patterns of fraud,¹⁸¹ increasing the expected value of suit. If plaintiffs can recover for each count but only have to pay experts to review a small portion of the relevant charts or case files,¹⁸² then profits would be greatest in large-scale Medicare fraud cases. Of course, at a certain point, widening the fraud's scope too much will make detection easier, suggesting a stronger chance that the government will step in. Nevertheless, the government's strained resources mean that it cannot prosecute many cases, particularly given the expense of prosecuting systematic fraud.¹⁸³

Even when plaintiffs could recover more than the cost of litigation, uncertainty reduces the expected value for qui tam plaintiffs significantly. One reason for this uncertainty is that the government can veto settlements for any reason, even when it has declined to intervene.¹⁸⁴ Recently, the Fourth Circuit

Richard Moberly, *Protecting Whistleblowers by Contract*, 79 U. COLO. L. REV. 975, 981 (2008) (“[A]nti-retaliation laws offer inconsistent relief to whistleblowers because of the wide variance in the scope of protections each provides.”).

180. See generally Loren H. Brown et al., *Bellwether Trial Selection in Multi-District Litigation: Empirical Evidence in Favor of Random Selection*, 47 AKRON L. REV. 663, 670-84 (2014) (explaining the process for bellwether trials).

181. This is true because standard errors are calculated based on overall number of observations, not percentage of total population included in the sample. See Tidmarsh, *supra* note 125, at 1468 (“Even if it does not increase the accuracy of individual awards, however, trial by statistics may do a better job of determining the aggregate liability of the defendant . . .”).

182. This depends, of course, on the cost of statistical experts. Since these experts may cost more per hour than medical experts, cost savings would depend on the number of hours needed to conduct the statistical analysis compared to the claim-by-claim examination. Whether sampling ultimately proves cheaper in practice is an empirical question that cannot be easily resolved *ex ante*.

183. See Topor, *supra* note 171 (“[A] lack of agency resources to pursue every worthy FCA action brought to its attention is frequently cited as a reason the DOJ might not intervene in an otherwise meritorious FCA case.”).

184. See 31 U.S.C. § 3730(b)(1) (“The action may be dismissed only if the court and the Attorney General give written consent to the dismissal and their reasons for consenting.”).

affirmed the Attorney General's right to veto any settlement for any reason whatsoever.¹⁸⁵ There is no reasonableness requirement in the FCA statute since, in effect, *qui tam* suits are brought on behalf of the government's interests, making the government's representatives the final arbiter of those interests.¹⁸⁶ While the legislative intent is fairly clear, there is additional uncertainty due to a circuit split on the government's ability to intervene at will in these suits, with the Ninth Circuit reading a reasonableness requirement into the statute.¹⁸⁷ This means there may be many *qui tam* plaintiffs for whom a settlement is otherwise profitable, but for whom uncertainty makes the necessary upfront investment unappealing.¹⁸⁸ While allowing sampling would not directly reduce this uncertainty, it could reduce the upfront costs associated with bringing a *qui tam* suit, and thus make trial a realistic option for plaintiffs without government backing.

While settlement may be uncertain because of the government's right to veto, relators also have an extremely strong incentive to settle so that they can avoid paying the cost of proving each and every fraudulent claim. Without sampling, defendants can force relators into settlements for the expected value of total damages minus the costs to take each claim to verdict. Since costs per claim are high, relators face significant pressure to accept nuisance-value settlements.¹⁸⁹ This high settlement pressure is largely an artifact of the high cost of proof and is larger in *qui tam* cases where the government does not intervene to shoulder some of the costs. Even though successful *qui tam* plaintiffs can recover reasonable costs,¹⁹⁰ costs are granted at the close of the case and only if the plaintiff prevails, which does little to resolve the issue of uncertainty dissuading upfront investment of capital. This settlement pressure encourages firms to continue fraudulent behavior because they will only pay a fraction of the actual harms they perpetrate against the government; in other words, the expected value of fraud against Medicare remains positive. Additionally, it may be that the expected value of recoverable damages for the

185. See *United States ex rel. Michaels v. Agape Senior Cmty., Inc.*, 848 F.3d 330, 339 (4th Cir. 2017) (“[T]he Attorney General possesses an absolute veto power over voluntary settlements in FCA *qui tam* actions.”).

186. See *id.* at 339-40.

187. See *United States ex rel. Killingsworth v. Northrop Corp.*, 25 F.3d 715, 722-23, 725 (9th Cir. 1994) (reading a reasonableness requirement into the Attorney General's authority to veto settlements).

188. See Tidmarsh, *supra* note 125, at 1467 (“[W]ithout a cost-effective method for deciding these cases, defendants could cause small-value harms on a large scale with impunity.”).

189. See *United States ex rel. Michaels v. Agape Senior Cmty., Inc.*, No. 0:12-3466-JFA, 2015 WL 3903675, at *2-3, *5 (D.S.C. June 25, 2015) (explaining that due to evidentiary costs, the plaintiffs had initially accepted a \$2.5 million settlement), *aff'd in part, appeal dismissed in part*, 848 F.3d 330.

190. See 31 U.S.C. § 3730(d)(2) (2017).

government is actually negative, preventing most plaintiffs from bringing suit in the first place. These twin dynamics allow defendants to profit from the inefficient administration of justice. Creating efficient settlement behavior is essential because settlement behavior determines the damages the vast majority of defendants ultimately pay.¹⁹¹

Two additional sources of uncertainty for qui tam plaintiffs militate against bringing suit: the sealed complaint rule and the first to file rule. The FCA requires relators to file complaints under seal for sixty days, until the government decides if it wants to intervene.¹⁹² Additionally, the right to act as a relator in a qui tam action with duplicate filings for the same false claims goes to the first person to file.¹⁹³ This could conceivably create a perverse situation where multiple complaints are filed under seal, only the government knows who has filed complaints, and all relators sustain the upfront costs to formulate a complaint, although only one relator has the chance to file. Of course, it is difficult to know for sure because this situation does not readily lend itself to empirical analysis. This uncertainty contributes to an environment of underdeterrence that makes reducing costs to file suit critical, although sampling would not directly resolve this issue.

The government has minimal incentive to intervene in FCA cases, primarily because of the high costs of intervention and the low reduction in recovery should the relators win without government intervention. This further undermines deterrence. If the government does not choose to intervene, it loses between 0% and 20% of its expected recovery.¹⁹⁴ In almost all cases, the cost of paying experts to code tens of thousands of charts will *vastly* exceed the earnings the government forgoes in choosing not to intervene.¹⁹⁵ Since the government has little incentive to intervene, almost all of the action will have to come from qui tam plaintiffs acting without government support. The combination of qui tam plaintiffs without support and government-initiated actions alone are likely to underdeter fraud, as discussed above. Reducing the costs of proof would increase the incentives for the government

191. See Tidmarsh, *supra* note 125, at 1460 (“Virtually every civil case settles, and class actions or other aggregate litigation are not exceptions to the rule. The real battle—the one that determines the value of the settlement—is whether the cases are aggregated in the first place.” (footnotes omitted)).

192. See 31 U.S.C. § 3730(b)(2)).

193. See *id.* § 3730(b)(5) (“When a person brings an action under this subsection, no person other than the Government may intervene or bring a related action based on the facts underlying the pending action.”).

194. See *id.* § 3730(d)(1)-(2).

195. See, e.g., *United States ex rel. Michaels v. Agape Senior Cmty., Inc.*, No. 0:12-3466-JFA, 2015 WL 3903675, at *2, *5 (D.S.C. June 25, 2015) (estimating 10% of damages to be \$2.5 million, and the cost to complete discovery at over \$16 million), *aff'd in part, appeal dismissed in part*, 848 F.3d 330 (4th Cir. 2017).

to intervene, further reducing the expected costs to qui tam plaintiffs. Freeing up budget space for additional government involvement would also make possible more frequent government intervention.

Finally, spoliation of evidence may make it very difficult or impossible to prove each individual claim, even if there is obvious evidence of false claims. For example, if there is strong evidence that a pattern of false shipping claims exists, but many previous shipments have been unpacked, it would be impossible to gather dispositive evidence about the content of *all* the shipments.¹⁹⁶ However, sampling to establish the pattern of fraud may provide a credible solution, reversing the perverse incentive for spoliation of evidence in false claims cases.¹⁹⁷

Ultimately, properly executed statistical extrapolation results in significant improvements in deterrence, since a precise, or nearly perfect, classification of claims incurring relatively low transaction costs greatly increases the likelihood that a plaintiff will step forward to file suit in the event of systematic Medicare fraud.¹⁹⁸

3. Negative impacts on government and relator interests

Allowing statistical extrapolation could have a chilling effect on efficient Medicare claims, creating concern that patterns of borderline claims may force providers into defending FCA suits. Nevertheless, overdeterrence is not costly in the Medicare context. For the most part, the only thing such a rule deters is creating a pattern of false claims, or at least claims of such dubious validity that an employee might blow the whistle on his or her employer. Such borderline fraud does not produce obvious social benefit, particularly in the context of Medicare, where many services are already condemned by health policy analysts for being unproductive or only marginally cost effective.¹⁹⁹ Given the

196. See, e.g., *id.* at *6 (“[T]his Court recently handled a . . . *qui tam* action where statistical sampling represented the only way the plaintiff-relators could prove damages. . . . [F]or the vast majority of the claims, the shipments had been completed and the belongings unpacked, thus rendering it impossible to determine if weight bumping had occurred.”).

197. In *Michaels*, the court suggested that the plaintiff should get the benefit of the doubt in the event that there were differences between the claims with spoiled evidence and those without. See *id.* at *6-7. The court determined, in effect, that the defendant should not benefit from its inequitable conduct. See *id.* at *7.

198. See Tidmarsh, *supra* note 125, at 1465-66 (“If the sampling is done well, the court obtains a good sense of the size of the harm that the defendant has caused, and can extrapolate from the sample to ensure that the defendant is held responsible for that amount of harm—no more and no less. Thus, defendants will internalize the costs of their behavior, and will be neither underdeterred nor overdeterred.”).

199. See, e.g., Jeffrey C. Brenner, *Proposal for a Demonstration Program for Accountable Care Organizations (ACOs) in Urban, Underserved Communities in New Jersey*, in AM. LAW INST., *footnote continued on next page*

overlap between unproductive medical spending and upcoding, it would not be surprising to find that very little questionable Medicare spending contributed to patient health in a noticeable way. In fact, one of the primary goals of the contemporary value-based care movement²⁰⁰ is reducing overutilization of Medicare services that produce limited marginal benefit at significant cost.²⁰¹ Overutilization is particularly acute for elderly populations, in which proper care management could result in significantly less utilization without a noticeable drop in quality.²⁰² Perhaps deterrence is actually socially beneficial from the perspective of care providers—limiting overutilization (and thus supply) constrains local area variation,²⁰³ as individual providers who want to offer more efficient care need not unilaterally break from their communities' norms for care provision and potentially lose customers.²⁰⁴ In addition, the perceived cost to potential relators of filing suit against a former employer discourages most nonmeritorious claims. To the extent that it does not,

NAVIGATING HEALTH CARE REFORM: CHALLENGES FOR INSURERS AND PROVIDERS 33, 34 (2010) (“The Dartmouth Atlas highlights unacceptable regional variations in cost and health care utilization for Medicare patients. It shows that costs in a state, region, city, or hospital are more tied to health care supply than patient need. . . . [High-cost regions] provide uncoordinated and often unnecessary services of no benefit to the patient.”).

200. “Value-based programs reward health care providers with incentive payments for the quality of care they give to people with Medicare.” *What Are the Value-Based Programs?*, CENTERS FOR MEDICARE & MEDICAID SERVICES, <https://perma.cc/2YHD-KL75> (last updated July 25, 2018).
201. See Brenner, *supra* note 199, at 37 (arguing that reforms to align the provision of Medicare services more closely with end results are crucial to ensure quality improvements).
202. See RISA LAVIZZO-MOUREY, ROBERT WOOD JOHNSON FOUND., *THE REVOLVING DOOR: A REPORT ON U.S. HOSPITAL READMISSIONS* 4 (2013), <https://perma.cc/X9XF-MAFK> (“[E]ven though hospitals are places where life-saving heroics are routine, they can also be costly and dangerous places to receive care.”).
203. Local area variation refers to large fluctuations between hospital service choices even within relatively confined geographical areas. See Yuting Zhang et al., *Comparing Local and Regional Variation in Healthcare Spending*, 367 *NEW ENG. J. MED.* 1724, 1725, 1727 (2012) (explaining the concept of local area variation and then finding a high level of variation even within small areas such as Manhattan).
204. See Elliot Fisher et al., Dartmouth Inst. for Health Policy & Clinical Practice, *Health Care Spending, Quality, and Outcomes: More Isn’t Always Better* 2 (2009), <https://perma.cc/5Z43-9ED5> (“Medicare beneficiaries in high-spending regions do not receive more ‘effective care’ Rather, the additional services provided to Medicare beneficiaries in higher-spending regions all fall into the category of ‘supply-sensitive care’: discretionary care that is provided more frequently when a population has a greater per capita supply of medical resources.”).

sanctions under Federal Rule of Civil Procedure 11,²⁰⁵ the plausibility pleading standard,²⁰⁶ and relatively strong application of summary judgment rules²⁰⁷ all discourage plaintiffs from filing weak claims.

D. Balancing

Courts must balance private interests, the risk of error in depriving private parties of their property, and government interests,²⁰⁸ all of which may be salient in the FCA context. As the U.S. Supreme Court explained in *Mathews*, “[a]ll that is necessary is that the procedures be tailored, in light of the decision to be made, . . . to insure that [parties] are given a meaningful opportunity to present their case.”²⁰⁹ Thus, the question is whether the use of sampling and forecasting in FCA claims can appropriately ensure that defendants will be able to meaningfully defend themselves while protecting the important public benefits that prosecuting large-scale frauds at lower cost would provide. Even if courts determine in the abstract that private interests in the accurate administration of each individual claim outweigh government interests in the efficient administration of justice, where variance between claims is small and thus statistical techniques are at their most effective, courts should allow statistical extrapolation in FCA cases alleging large patterns of fraud.

Courts balancing private interests, the risk of erroneous deprivation, and government or relator interests should err heavily in favor of the last two factors.²¹⁰ First, due process precedent supports the elevation of significant efficiency gains above minor risks of error or misclassifications of liability. For example, the court in *In re Estate of Marcos* observed that the risk of underdeterrence significantly outweighed any minor risk of overcompensa-

205. See, e.g., *Pentagen Techs. Int’l Ltd. v. United States*, 172 F. Supp. 2d 464, 471-73 (S.D.N.Y. 2001).

206. See *Ashcroft v. Iqbal*, 556 U.S. 662, 678 (2009) (“To survive a motion to dismiss, a complaint must contain sufficient factual matter, accepted as true, to ‘state a claim to relief that is plausible on its face.’” (quoting *Bell Atl. Corp. v. Twombly*, 550 U.S. 544, 570 (2007))); *Twombly*, 550 U.S. at 554-57 (holding that complaints must state sufficient nonconclusory allegations to rise to the level of plausibility).

207. See *Celotex Corp. v. Catrett*, 477 U.S. 317, 325-27 (1986); see also FED. R. CIV. P. 56.

208. See *Mathews v. Eldridge*, 424 U.S. 319, 334-35 (1976).

209. *Id.* at 349.

210. The government’s interest includes efficiency and prompt resolution. See Laurens Walker & John Monahan, Essay, *Sampling Liability*, 85 VA. L. REV. 329, 343 (1999) (“Individualized information should be used where it is practical—i.e., cost effective—to obtain. If individual information is not practical to obtain, however, sampling should be used so that a judgment can be reached efficiently and expeditiously.”).

tion created by statistical extrapolation.²¹¹ Given the lack of clinical value in most borderline Medicare claims, prioritizing fraud prevention also seems reasonable from a health care perspective.²¹² Second, *Daubert* hearings and strict application of Federal Rule of Evidence 702 procedures to evidence of statistical extrapolation already protect defendants' rights in other contexts, such as mass torts, antitrust suits, and other complex litigation.²¹³ It is unclear why the risk of error is larger in this context compared with class actions alleging violations in other areas of the law; instead of variation in *both* parties and the particulars of claims, the FCA involves variation only in specifics of claims. Thus, the case for allowing sampling to demonstrate liability in large-scale FCA cases appears sound. Courts should heed *Martin* and grant permission to utilize such techniques—subject to significant pretrial safeguards, discussed in the next Part.

IV. Guidelines for Implementing Statistical Analysis

Even if courts follow *Martin*, they will need to design specific safeguards to implement extrapolation techniques without violating due process, particularly given the mathematical complexity of determining the risk of erroneous deprivation. This Part analyzes several options and addresses the issues with potential alternatives to statistical extrapolation, particularly the bellwether trial. Next, it suggests a framework for determining when sampling may be used, drawing from experience with class certification procedures. Finally, it explains that even when courts have proven generally resistant to sampling to prove liability, they have still recognized the validity of specific exceptions.

A. Modified Bellwether Option

The most commonly proposed solution to an overabundance of claims, and the likely next-best alternative to allowing sampling to prove liability, is

211. See *In re Estate of Marcos Human Rights Litig.*, 910 F. Supp. 1460, 1468 (D. Haw. 1995) (“The judicial and administrative time and costs of holding bipolar trials would also have been virtually, if not absolutely, prohibitive. Lastly, because class members are mostly impecunious, the cost of bringing them to the forum or even taking their depositions would have prevented their claims from ever being determined.”), *aff’d sub nom.* *Hilao v. Estate of Marcos*, 103 F.3d 767 (9th Cir. 1996).

212. See generally Richard Smith, “Flat of the Curve” Healthcare, BMJ OPINION (Mar. 23, 2015), <https://perma.cc/2JV6-VSJ6> (describing the low value added from marginal health care expenditures after a certain level of per person spending).

213. See Rhoad et al., *supra* note 94, at 1-3 (explaining how defendants might use such protections in the FCA context).

the bellwether trial.²¹⁴ Bellwether trials get their name from the practice of placing a bell on a lead “wether,” or male sheep, to lead the movements of a flock of sheep.²¹⁵ The legal context is similar: Courts take a subset of claims and try them in order to give guidance for settlement negotiations over the remaining pool of claims.²¹⁶ While this option has intuitive appeal, it does not provide sufficient guidance for jurors, and would replicate the worst parts of statistical extrapolation, discussed below, without capturing any of its benefits.

Proposals to date incorporating the bellwether procedure into FCA cases advocate duplicative processes with little added benefit. For example, Milene Vega has suggested proving liability in a bellwether trial.²¹⁷ Her proposal appears to limit these trials to the question whether a systematic pattern of fraud exists.²¹⁸ There is no reason to submit this question to a jury, because it is largely a question of evidence reliability: If the variance of the entire population of cases is too large, then extrapolation will produce increasingly erratic results.²¹⁹ Putting questions about the reliability of expert evidence before a jury may produce unreliable results, which is precisely why judges perform their crucial gatekeeping function under the Federal Rules of Evidence.²²⁰ The proposal also confuses the relevant statistical question.²²¹ There need not be an *intended* or *knowing* pattern of fraudulent conduct for sampling to reliably estimate liability.²²² Instead, the relevant inquiry is

214. See Vega, *supra* note 24, at 589 (arguing that bellwether trials can be used to determine whether there is “sufficient evidence of a generalized policy of fraud” before using statistical sampling).

215. See *In re Chevron U.S.A., Inc.*, 109 F.3d 1016, 1019 (5th Cir. 1997).

216. See *id.*; see also *United States ex rel. Michaels v. Agape Senior Cmty., Inc.*, No. 0:12-3466-JFA, 2015 WL 3903675, at *2 (D.S.C. June 25, 2015) (discussing doing precisely this, but ultimately canceling the bellwether trial because of a settlement), *aff’d in part, appeal dismissed in part*, 848 F.3d 330 (4th Cir. 2017).

217. See Vega, *supra* note 24, at 589.

218. See *id.* at 589-90.

219. See *supra* notes 164-66 and accompanying text.

220. See *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579, 592-93, 597 (1993); see also FED. R. EVID. 702.

221. For example, Milene Vega’s distrust of sampling is evident in her claim that disallowing sampling “would . . . focus [the government’s] efforts on the claims it can actually prove.” See Vega, *supra* note 24, at 588. This overemphasis on qualitative proof mistakes a lack of rigorously bounded error for objective veracity. The precision of “actual proof” would be greatly increased if the errors in assigning liability could be weighed in an objective sense, which is precisely what modern econometric techniques are designed to do. See *supra* Part I.B.

222. *But see* Vega, *supra* note 24, at 589-90 (“I suggest that the use of statistical sampling to extrapolate liability in FCA cases concerning healthcare fraud should be limited to cases where evidence is presented (again prior to the use of statistical sampling) indicating a known company-wide procedure or policy that results in the systematic submission of false claims.”).

whether there is sufficient *statistical* connection between the cases, such that extrapolation is highly likely to produce accurate and precise categorization of liability. Even if this were Vega's intent, submitting the question whether a pattern of fraud exists to a lay panel would likely collapse the inquiry into the less complicated question of culpable mental state, simply because the statistical analysis would be too difficult.²²³ This would both be duplicative and eliminate the error-reducing benefits of extrapolation.

If the bellwether plaintiffs are selected to accurately and precisely represent the underlying population of claims, then bellwether trials are functionally indistinguishable from the option of selecting a representative sample,²²⁴ except that the defendant retains the right to threaten to try all remaining counts individually. This is tantamount to certifying a class, but allows the defendant to demand trials for each individual plaintiff after receiving an unfavorable decision against the named plaintiff. Not only does this create an administrative headache with little discernible benefit, it protects a right not in need of protection.²²⁵ Thus, there is little additional benefit in terms of the *Mathews* balancing from adopting a court-selected bellwether trial approach; instead, the bellwether approach simply creates additional procedural hurdles for the government or relators, once again raising the cost of prosecuting FCA claims and contributing to underdeterrence. In other words, such an approach would obviate the advantages to extrapolation outlined above.²²⁶

The idea that each individual FCA claim is unique, and therefore in need of individualized proof, does not reflect the reality of many patterns of fraud. Just as many plaintiffs may have sufficient similarities to justify trying their claims together (whether through joinder or class action), many claims are sufficiently similar conceptually to make the exercise of individualized proof redundant and inefficient.²²⁷ Human behavior is often far more systematic than many are willing to accept; patterns of fraudulent claims are likely to be

223. *But see id.* (suggesting that judges and juries could properly select between these alternatives). Vega also suggests that the interests of "justice" outweigh any potential benefits in terms of judicial economy. *See id.* In my view, she does not place the question into sufficient context, and her conclusion is too dismissive. *See supra* Part III (conducting a more extensive cataloguing of the factors involved in the *Mathews* balancing).

224. *See* Amir Seyedfarshi, *Binding Bellwether Trials in Multidistrict Litigation and the Right to Jury Trial*, 17 W. MICH. U. COOLEY J. PRAC. & CLINICAL L. 295, 319-20 (2016) (laying out the sampling procedure used to construct a representative bellwether sample).

225. *See supra* Part III.D.

226. *See supra* Parts III.C.1-2.

227. *See* Tidmarsh, *supra* note 125, at 1468 ("[E]ven if it does not increase the accuracy of individual awards, however, trial by statistics may do a better job of determining the aggregate liability of the defendant." (footnote omitted)).

easily discernible, especially in large samples of claims.²²⁸ For example, the defendants in *Martin* objected to statistical sampling to establish liability, claiming that individualized factors determined the proper course of care, such as “age; gender; pre-hospital condition/prior level of function; reason for hospitalization; [and] condition upon admission to [a] skilled nursing facility.”²²⁹ The court correctly pointed to the fallacy of this reasoning:

Defendant’s argument highlights the very nature of statistical sampling: that a smaller portion of claims will be used to draw an inference about a larger, not entirely identical, population of claims. . . . If all of the claims were exactly the same in every respect, there would be no need for statistical sampling and extrapolation in litigation because each individual unit would be identical . . .²³⁰

Nevertheless, if courts remain skeptical of wholesale adoption of statistical extrapolation, requiring appointment of a representative to select bellwether plaintiffs represents a second-best option.²³¹ A properly conducted bellwether trial should still decrease pressure on plaintiffs to settle,²³² increase incentives for defendants to internalize costs, and reduce the total costs of bringing suit for both parties.²³³ In other words, bellwether trials achieve some limited benefits of extrapolation, but still give too much power to defendants, and this cannot wholly rectify the perverse incentives the current FCA scheme creates for plaintiffs.

B. Class Action Analog

Rather than attempt to use a new version of the existing bellwether procedure, courts could draw from their experience in class action cases to fashion a set of rules for increasing the likelihood that extrapolation will produce reliable results in FCA cases. This Subpart argues that courts should

228. See GREENE, *supra* note 165, at 62.

229. See *United States ex rel. Martin v. Life Care Ctrs. of Am., Inc.*, 114 F. Supp. 549, 566 (E.D. Tenn. 2014).

230. *Id.*

231. A bellwether trial may ultimately be easier for a court to administer because it would not require complex proof of the rate of misclassification during the pretrial hearing. Instead, the court could simply use a random number generator to draw a selection of claims, and allow trial to proceed on those claims.

232. Although this is true, plaintiffs would still have a major incentive to settle for less than their claims are really worth, because the defendant retains the right to force the plaintiff into trials on the remainder of the claims. The threat of imposing costs on the plaintiff gives the defendant significant settlement leverage that would not exist if all claims were tried at the same time using the sampling method.

233. See *Brown et al.*, *supra* note 180, at 667 (explaining the efficiency benefits from conducting bellwether trials, as well as the particular places where costs may be reduced).

require a large number of claims, strong similarity between claims, and adequacy of the representativeness of the plaintiff for the government's interests in the case.²³⁴

The existing Federal Rule of Civil Procedure governing class action certification provides guidance for aggregation processes that comport with due process.²³⁵ Establishing numerosity, adequacy, and commonality thresholds in the FCA context would alleviate due process concerns.²³⁶ Requiring that common questions of law and fact predominate over questions affecting individual claims, or that adjudications with respect to individual claims dispose of material issues in the other claims, would capture the spirit of the requirements of Rule 23(b).²³⁷ However, this would require breaking with class certification in an important way: The factors would apply to claims, rather than parties. While normally this would be done through the joinder process, there is no compelling reason the same set of features could not apply to a single party with an unmanageably large set of claims; it already applies in the class action context, in which both the party *and* the claims vary.²³⁸ In fact, there is no class certification requirement that deals with the identity of the party as such—only the typicality and adequacy of the named party's claims with respect to the broader class of claims.²³⁹

Unfortunately, this concept requires broadening the already unpopular class action rules and applying them to a new set of cases, making it potentially unpalatable.²⁴⁰ While concern for the welfare of corporations is ever popular in Washington, protecting corporations that defraud Medicare may strike many politicians as a less than savvy move, as the public is unlikely to accept

234. Although these requirements mirror class action requirements, this is not a suggestion to literally amend Rule 23.

235. See FED. R. CIV. P. 23(a).

236. While Rule 23(a) also requires that a representative plaintiff present claims typical of the class, *see id.*, the representativeness requirement is trivial in the FCA context because there is only a single plaintiff.

237. See *id.* r. 23(b).

238. See *Tyson Foods, Inc. v. Bouaphakeo*, 136 S. Ct. 1036, 1045-46 (2016) (explaining that class actions allow for consolidation of multiple causes of action by multiple parties, even when all claims and parties are not identical, as long as common issues predominate).

239. See FED. R. CIV. P. 23(a)(3)-(4).

240. See Ese Olumhense, *Republicans Introduce Bill to "Kill" Class Action Lawsuits*, TRIB. MEDIA WIRE (updated Mar. 9, 2017, 5:05 PM), <https://perma.cc/H37X-VETC> (describing Republican plans for legislation to reduce class action lawsuits); *see also* Fairness in Class Action Litigation and Furthering Asbestos Claim Transparency Act of 2017, H.R. 985, 115th Cong.

such a position. While this consideration does not enter explicitly into the *Mathews* framework, judges are likely to consider this backdrop when determining whether to allow sampling.

Another potential cost of this system is that it may invert the existing settlement pressure on plaintiffs by placing too much pressure on defendants.²⁴¹ On the other hand, defendants would also save on defense costs, since they would not have to screen every claim in preparation for trial. To the extent that this creates a chilling effect, it may increase the incentives for care providers to offer efficient quantities of care.²⁴² In other words, creating settlement pressure on defendants may align government and private interests in efficiently preventing Medicare fraud. The risk of erroneously depriving defendants of profits comes almost entirely from the prospect of encouraging strike suits,²⁴³ but clear filing guidelines like those governing class actions could at least defray this risk. Of course, the clearer the courts' guidance, the less risk of creating undue pressure on defendants to settle strike suits.

Rather than using all of the class certification rules, courts could adopt an abbreviated test, highlighting numerosity as a proxy for administrability. This would permit sampling in particularly unwieldy case without broadly allowing its use to prove liability. There is some flexibility in this concept, since it potentially justifies both broader and narrower versions of the existing class certification rules. It could be broader if it simply eliminates the other requirements. Since commonality functionally equates to requiring low variance in the population of claims, it becomes a relevant constraint when there are large numbers of variable claims. Nevertheless, this administrability rule may also restrict use of sampling, particularly if courts require more rigorous proof of difficulty in administering claims through joinder than in the class context. In more concrete terms, class certification typically only requires roughly thirty plaintiffs before numerosity is fulfilled; administrability here implies thousands, or tens of thousands, of claims.²⁴⁴

241. See Richard O. Faulk, *Armageddon Through Aggregation?: The Use and Abuse of Class Actions in International Dispute Resolution*, 10 MICH. ST. U. DET. C.L. J. INT'L L. 205, 205 (2001) ("The system of justice . . . which guarantees individual plaintiffs and defendants their 'day in court[]' is increasingly being sidestepped by procedural rules that allow entrepreneurial lawyers to aggregate claims into massive controversies that . . . [impose] enormous and intolerable risks which defendants cannot prudently accept . . .").

242. See *supra* Part III.C.

243. Strike suits are suits that have marginal expected recovery, but which encourage defendants to settle for nuisance value rather than litigate.

244. See *United States ex rel. Michaels v. Agape Senior Cmty., Inc.*, No. 0:12-3466-JFA, 2015 WL 3903675, at *1 (D.S.C. June 25, 2015) (explaining that because there were around 53,000 to 62,000 claims, it would be difficult to try each individually), *aff'd in part, appeal dismissed in part*, 848 F.3d 330 (4th Cir. 2017).

Typical FCA claims would not qualify; only serial abusers of government funding need worry.²⁴⁵ Such cases are relatively rare, and also require the strongest deterrent.²⁴⁶ This rule may produce the greatest gains with the least perturbation to the status quo, given that serial frauds present the hardest claims to tackle under the existing system, since the parties' incentives are most misaligned.²⁴⁷

Overall, the best solution would draw upon experience with class action concepts without relying on the flawed bellwether trial formula to determine where sampling is appropriate. This would avoid many of the troubles of class certification, because the plaintiff is only a single party rather than an actual class of people. Finally, it would better align plaintiffs' incentives to bring cases with the FCA's goal of reducing fraud.

Conclusion

Fixing ever-expanding Medicare budget outlays requires finding new and more efficient ways to encourage high-quality care and deter fraud. Recent rulings on the validity of statistical sampling to prove liability for Medicare fraud under the FCA create a unique opportunity to revisit the application of "trial by statistics" outside of the class action context. The U.S. Supreme Court's recent guidance in *Tyson Foods, Inc. v. Bouaphakeo* breathed new life into this question, creating a window for courts to reconsider the due process issues implicated by the use of extrapolation to prove liability. Courts concerned about due process challenges should look at the bigger picture: Fraud against the government is woefully underprosecuted, qui tam plaintiffs have little incentive to file suit, and defendants have a perverse incentive to increase the scope of their fraudulent conduct in order to dissuade suit. This is a broken system with an easy solution: allowing statistical extrapolation for proof of liability in FCA cases—along with sufficient safeguards, similar to *Daubert* hearings or motions to certify a class. The justice system's faith in the fiction of granting every defendant its "day in court" has gone too far; systematic fraud demands admission of systematic statistical evidence.

245. Even then, they need not worry very much, because the government simply does not have the resources to prosecute a significant number of FCA claims. See U.S. GOV'T ACCOUNTABILITY OFFICE, GAO-13-283, HIGH-RISK SERIES: AN UPDATE 246-59 (2013).

246. See *supra* Part III.C.2.

247. See *supra* Part III.C.