Blog Post

AI-Assisted Data Mining and the False Claims Act: Is a Data Anomaly Really a False Claim?

September 2, 2026

By Noam B. Fischman and Ameer Al-Khudari

Data miner whistleblowers, increasingly armed with artificial intelligence (AI) tools, are already reshaping how False Claims Act (FCA) theories are built, generated, and litigated. The Department of Justice's (DOJ) April 2026 announcement of the Fraud Oversight through Careful Use of Statistics (FOCUS) initiative, which will strengthen relationships with data miner relators who have developed "effective tools," only solidified what the False Claims Act bar had already portended. Not surprisingly, the commentary that followed focused on the public disclosure bar as a notable defense. But what about scienter as a defense? And what does the *subjective* scienter analysis mean for a relator that is a data miner, an industry outsider whose allegations come from a publicly available dataset, not a seat at any corporate table or a desk down the hall from the allegedly bad actors? In a post-FOCUS world, emphasizing the subjective intent standard will become a critical defense strategy.

Data Mining and the FCA Landscape

Data miners are not traditional whistleblowers with insider knowledge. They are analysts — often AI-assisted — who mine public datasets for statistical outliers and then file *qui tam* complaints based on publicly available government data. On April 30,

Related People

Ameer Al-Khudari
Noam B. Fischman

Related Work

Healthcare
Healthcare Fraud and Abuse
Healthcare Litigation

Related Offices

Chicago
Washington, D.C.

Health Law Rx

Akerman Perspectives on the Latest Developments in Healthcare Law

Read blog posts

2026, DOJ announced FOCUS to formalize its relationship with those data miners. The timing is notable: data miners have accounted for more than 45% of all qui tam complaints since FY 2024. Filings surged to 980 in FY 2024 and nearly 1,300 in FY 2025; more than 780 were filed through mid-year FY 2026, with another record year in sight. DOJ has said that FOCUS would prioritize data miners who demonstrate pre-filing diligence, analytical rigor, familiarity with program rules, and legally sufficient allegations.

Data-miner relators typically use public datasets — Medicare and Medicaid claims data published by CMS, SBA PPP loan disclosures, customs and trade data, grant databases, contract award databases, state corporate registries, and similar sources — to identify statistical outliers. Their tools crunch the data, cross-reference it against benchmarks, and flag patterns for follow-up. The findings may flag a provider, for instance, whose billings diverge from peers, a PPP borrower whose loan amount appears inconsistent with public business data, or a contractor whose pricing looks unusual against a benchmark.

Not surprisingly, the response of the defense bar has been unequivocal: a statistical outlier does not constitute fraud. Legitimate business variations occur as do data-entry errors, differences in service mix, geographic factors, or other innocent explanations. Technology may sharpen the signal, but it does not close the gap between an unusual data point and a false claim submitted with the requisite state of mind.

Defense No. 1

Public Disclosure Bar and DOJ's Veto Power

The public disclosure bar is therefore an obvious first-line defense. Briefly, pursuant to 31 U.S.C. § 3730(e)(4)(A), a court must dismiss a *qui* *tam* action — unless the government opposes dismissal — if substantially the same allegations or transactions were publicly disclosed through specified channels. The exception to this dismissal imperative is that a relator must qualify as an "original source" of the information giving rise to the claim(s).

Public disclosure/original source discussion presents a dilemma for data miners. Definitionally, data miners look at publicly available data, which is the quintessence of what the public disclosure bar was meant to prevent. Necessarily, the question pivots to whether a data miner can be an original source — *i.e.*, someone who has voluntarily disclosed information to the government before the public disclosure or someone who has knowledge that is independent of and materially adds to the publicly disclosed allegations or transactions.

The law is relatively steady on this point. A relator materially adds to public disclosures only if her information is "'sufficiently significant or essential' to influence the government's decision to prosecute."[1] Merely adding "detail or color" to a publicly disclosed scheme does not qualify.[2] By contrast, specific, non-public evidence — for example, firsthand knowledge that a designated entity had no office, no employees, no website, and no working phone number — can meet the "materially adds" threshold.[3]

Nuance matters here. Generally speaking, a data miner is running algorithms on CMS claims data or SBA loan disclosures. They apply a scientific methodology. Without more, that data miner does not contribute the sort of nuanced, non-public, entity-specific evidence that courts have found sufficient.

Courts are still working through how the public disclosure bar applies to data-mined theories. Different decisions reflect differing approaches across jurisdictions. Some courts have found public databases sufficient to trigger the bar when a relator's theory arises solely from those public sources.[4] Others (as recently as last month) have required a closer match between the database and the relator's allegations.[5]

Critically, although the public disclosure bar is an obvious first-line defense, the statute gives the relator a "get out of jail free" card insofar as the government can veto a dismissal on this specific ground. So, what's next?

What's next?

Defense No. 2: cue the rise of *Schutte*

The Supreme Court's decision in *United States ex rel. Schutte v. SuperValu Inc.*, 598 U.S. 739 (2023) (*Schutte*) was widely viewed as a loss for FCA defendants because it removed a tool from a defendant's toolkit: reliance on an objectively reasonable reading of a statute or regulation would no longer be the skeleton key to an early dismissal of a FCA claim. Instead, *Schutte* augured in a new world in which the FCA's scienter analysis turned on the defendant's *subjective* knowledge or belief at the time of the claim submission.

Counsel (undersigned as well) authored countless blogs that foretold the death of the scienter defense for many FCA cases. Were we wrong?

Data miners have an external view of a healthcare company only. The process itself is misaligned with what *Schutte* requires. Data miners have no access to internal emails, compliance warnings, conversations with supervisors, or other evidence of what the defendant actually knew or actually believed. A data miner cannot provide necessary color — who said what to whom and when? Did anyone ignore (or override) compliance functions? What did the billing folks actually think/intend? Even at the initial pleading stage, when scienter is typically governed by *Iqbal* rather than rule 9(b), corporate outsiders may still be unable to plead anything beyond conclusory tropes regarding scienter that do not suffice.

The relator's likely fallback argument will be to infer scienter from the size of the discrepancy: the pattern is so stark, the argument goes, that the defendant ***must*** have known. But that rhetorical structure creates a weight so heavy that it could collapse the critical distinction between the current subjective (*Schutte*) and previous objective reasonableness (*pre-Schutte*) scienter standards. An outside observer's assessment of the magnitude of an issue is no substitute for allegations of the defendant's contemporaneous knowledge or belief, which is now what the Supreme Court requires. That problem (for the relator) only compounds when the pattern was identified with AI rather than a human who also never knew or functioned within a defendant's individual ecosystem. Put another way, automation can make the signal more precise, but it cannot show that the defendant saw the pattern, understood its significance, or believed its claims were false.

Conclusion

Data miners using AI can find an anomaly. They can compile, compare, rank, and visualize public information at a scale no human team could match. But the FCA is a fraud statute, not an anomaly statute. It requires falsity, scienter, materiality, and more — and *Schutte* makes the defendant's subjective knowledge or belief central to that scienter inquiry.

The public disclosure bar and DOJ's position remain important first-line issues, but they are only the setup. The harder and potentially more consequential question is whether a data miner can connect an AI-detected pattern to evidence of what the defendant knew or believed when it submitted the claim.

The best scienter defense starts before the complaint arrives. Defendants who preserve contemporaneous compliance records, program interpretations, counsel reliance, and decision-maker rationale will be better positioned to show what they actually knew — not what an algorithm later inferred. Akerman has a specialized team of attorneys dedicated to advising and guiding clients with healthcare compliance as well as FCA, Anti-Kickback, and Stark Law investigations and litigation.

---

[1] *United States ex rel. O'Connor v. USCC Wireless Inv., Inc.*, 128 F.4th 276, 289 (D.C. Cir. 2025) (quoting *United States ex rel. Winkelman v. CVS Caremark Corp.*, 827 F.3d 201, 211 (1st Cir. 2016)); *United States ex rel. Reed v. KeyPoint Gov't Sols.*, 923 F.3d 729, 755-56 (10th Cir. 2019) (adopting *Winkelman*); *United States ex rel. Maur v. Hage-Korban*, 981 F.3d 516, 527 (6th Cir. 2020) (same).

[2] *Winkelman*, 827 F.3d at 211.

[3] *United States ex rel. O'Connor v. U.S. Cellular Corp.*, 153 F.4th 1272, 1281-82 (D.C. Cir. 2025).

[4] *United States ex rel. Stebbins v. Maraposa Surgical Inc.*, No. 24-1626, 2024 WL 4947274 (3d Cir. Dec. 3, 2024) (unpublished, non-precedential); *Relator LLC v. Cardinalli*, 2025 WL 4036760 (C.D. Cal. Sept. 15, 2025).

[5] *Relator LLC v. Erskine*, 182 F.4th 785 (9th Cir. 2026).

This information is intended to inform firm clients and friends about legal developments, including recent decisions of various courts and administrative bodies. Nothing in this Practice Update should be construed as legal advice or a legal opinion, and readers should not act upon the information contained in this Practice Update without seeking the advice of legal counsel. Prior results do not guarantee a similar outcome.