Cleared Does Not Mean Examined

Why a compliance screening program can pass every audit and still not detect the thing it exists to catch.

PUBLISHED	STATUS	REVISED
9 AUGUST 2026	PUBLISHED	9 AUGUST 2026
VERSION	**AUTHOR**	**LICENSE**
1.0	COLLIN B. GEORGE, CISSP	CC BY 4.0

Analytic record

TYPE	Analysis — a structural account of why counterparty screening programs measure their own activity rather than their effectiveness, and the design change that corrects it. Stable except where the underlying legal authorities change.
GOVERNING JUDGMENT	A screening program learns only from the counterparties it flags; the ones it clears are cleared without inquiry and teach it nothing. Its performance numbers are therefore conditional on its own past decisions, and a clean record on a class of counterparty is evidence the class was never examined — not evidence it is clean. The correction is a random-sample audit of cleared counterparties, which is inexpensive and is a design change rather than a purchase.
PROBABILITY	That screening validated only by process metrics cannot estimate its own failure rate is assessed as certain; it follows from the structure of the data. That attributes cheap for a counterparty to change carry most screening weight in typical programs is assessed as almost certain, from the enforcement record and standard program design.
ANALYTIC CONFIDENCE	High for the legal analysis, which rests on Supreme Court and Circuit authority and on Department of Justice settlement records. High for the structural argument, which restates an established result from program evaluation applied to a domain that has not absorbed it. These are separate axes and are not combined.
SUPERSESSION	Reassess on any Supreme Court decision altering False Claims Act materiality or scienter, on finalization of the CMMC assessment regime, or on new Civil Cyber-Fraud Initiative resolutions that change the enforcement pattern described here.
REVISION HISTORY	1.0, 9 August 2026 — first publication. Companion academic paper published concurrently on SSRN.

Executive summary

— Almost every screening program measures activity: transactions screened, watchlist currency, alerts cleared, clearance speed. None of those numbers answers the only question an enforcement authority asks — did the program catch the conduct it exists to catch.

— The gap is structural, not a reporting oversight. A screening program learns only from the counterparties it flags. The ones it clears are cleared without inquiry, so a clean record on a class of counterparty means the program never looked, not that the class is clean.

— The screening that happens leans on the wrong attributes: the ones a sophisticated counterparty can change for nothing — a corporate name, a registered address, a declared end-use, a self-attested compliance score. The attributes that are expensive to fake mostly go unscreened.

— The fix is one inexpensive design change: sample your own cleared counterparties at random and examine them, the way the IRS estimates the tax gap. That single stratum yields the one number no process metric can — an unbiased estimate of what your screening missed.

— This is now a False Claims Act question. A self-attested compliance score is a certification, and a false certification is a claim. Enforcement has already turned a self-reported score of 110, against a government assessment of negative 170, into a paid settlement — with no breach, no attack, and no data loss.

IF YOU READ NOTHING ELSE

Ask your screening program one question it probably cannot answer: among the counterparties we cleared, what fraction carried disqualifying conduct we missed? If the honest answer is that there is no way to know, the program is measuring its own activity and not its effectiveness — and it cannot substantiate a representation of reasonable care if one is ever tested. The correction is a random-sample audit of cleared counterparties. It is a design change, not a software purchase, and it is described below.

A question sits underneath almost every compliance program that screens counterparties — vendors, customers, distributors, end-users, suppliers: how do we know the screening works? The instinct is to point at the program's output. We screen every transaction. Our lists are current. Our alerts are cleared within a day. Those are real numbers and they are worth having. But none of them answers the question that was asked, and the reason is structural.

What a Screening Program Actually Learns From

A screening program is trained, formally through a model or informally through analyst experience and rule changes, on its own history. The counterparties that generated alerts were investigated. Those investigations produced outcomes, and the outcomes refined the rules. The counterparties that cleared were cleared without inquiry and produced no outcome at all.

That asymmetry is the whole problem. The program has evidence only about the population it already suspected. About the population it cleared — which is nearly all of it — it has nothing, because it never asked. Every performance number the program reports is therefore conditional on its own past decisions: the hit rate, the false-positive rate, the clearance accuracy all describe the slice of the world the program chose to look at.

The practical consequence is that a clean record on a class of counterparty is not evidence the class is clean. It is evidence the program never examined it. A profile that has never triggered an alert looks safe for exactly the same reason an unopened box looks empty. This is why compliance programs are

repeatedly surprised by enforcement actions involving counterparties they had screened and cleared many times over.

> **The clearance was never a finding. It was the absence of a question.**

> **WHY THIS MATTERS**
>
> A program cannot state its own failure rate if it has only ever examined the counterparties it flagged. That number — the rate of disqualifying conduct among the counterparties you cleared — is the one an examiner, an opposing expert, or a whistleblower will eventually put to you, and process metrics cannot produce it.

The Capability Nobody Measures

There is a second structural limit, and it sits upstream of everything the screening rules do. A screen is only as good as its ability to recognize that two nominally distinct counterparties are the same operation.

Evasion networks, sanctions front companies, and diversion schemes change corporate identity far faster than they change the substrate underneath it — the people, the physical premises, the freight forwarders, the banking relationships, the registered agents. A program that treats each new corporate name as a new and unknown entity discards the single most stable signal available to it, and re-clears an operation it has already seen under a different name.

Every screening program performs this entity resolution, whether it knows it or not. Almost none measure how well. The resolution quality is invisible to program management and to auditors alike, which means the binding constraint on the program's accuracy is a number nobody reports. A program that cannot say what fraction of its screened counterparties it has resolved to a persistent identity does not know how it performs against the counterparties most likely to be hiding.

> **WHY THIS MATTERS**
>
> Ownership-based rules — the OFAC 50 percent rule, foreign-ownership-and-control determinations, research-security affiliation checks — are graph problems. A program running them without resolving entities is performing the analysis it is able to perform, not the analysis the rule requires.

The Attributes You Trust Most Are the Ones an Adversary Can Change for Free

Screening weight tends to concentrate on the attributes that are cheapest for a sophisticated counterparty to alter.

A corporate name changes in days for a nominal fee. A registered address changes as fast. A declared end-use or end-user statement costs nothing to write. A self-classification, a declared value, a self-

attested compliance score — all are supplied by the counterparty and cost nothing to misstate. Against those sit the attributes that are genuinely expensive to fake: an operating history, a verified physical footprint, relationship tenure with banks and brokers and forwarders, a transaction pattern consistent with the stated business. Those mostly go unscreened.

> **An attribute's value to a screen is not its predictive power. It is its predictive power discounted by the cost to the counterparty of falsifying it.**

A feature that predicts well and is free to change is a liability, not an asset: it produces accuracy against unsophisticated actors, who never bother to change it, and close to none against the counterparties who will — while generating documentation that the program appeared to be working. Denied-party list screening is the pure case. It is highly accurate against entities that have not changed their names, and structurally uninformative against those that have.

> **WHY THIS MATTERS**
>
> A program optimized on historical accuracy alone will load its weight onto the cheap attributes, because the enforcement record it learned from is dominated by the unsophisticated actors who left them unchanged. It ends up optimized against the population least likely to cause a serious loss.

The Fix Is One Inexpensive Design Change

The correction to all three defects is a single addition, and it is cheap.

Sample your own clearances. Take a small fraction of the counterparties your screening cleared — one to three percent is enough at commercial volume — selected at random, independent of any score or alert or analyst judgment, and examine them properly. Stratify the sample by jurisdiction, by commodity or service class, by counterparty tenure, so each part of your book is represented.

This is not a novel technique. It is exactly what the Internal Revenue Service does to estimate the tax gap: a randomly selected audit stratum, chosen independently of the enforcement-selection rules, precisely so it reveals what routine enforcement misses. The design has been in use for decades in a domain with far higher stakes than most compliance programs face.

That single stratum produces the one figure no process metric can: an unbiased estimate of your clearance failure rate, with a confidence interval. It lets you state, in an audit or a courtroom, that among the counterparties your program cleared, the estimated rate of undetected disqualifying conduct is some measured value. Without it you can say nothing at all about the counterparties you cleared — which is nearly all of them. The objection to the sample is never its cost, which is trivial. The objection is that it generates a documented estimate of the program's own failure rate. That is the argument for it, not against it, and the reason is legal, addressed below.

Around that core, four supporting moves follow directly from the three defects:

- **Weight attributes by how hard they are to fake**, and cap the weight assigned to cheap-to-change attributes even where they predict well. Accept slightly lower measured accuracy in exchange for accuracy against the counterparties who matter.
- **Resolve counterparties into a persistent entity graph**, so a re-registered shell is recognized rather than cleared as new, and report the resolution rate as a program metric.
- **Let trusted status decay.** Any status that lowers scrutiny — approved end-user, certified supplier, an accepted remediation plan, a prior favorable review — should carry an expiry and be suspended by behavioral triggers, not left standing until a formal revocation. The status a program trusts most is the most valuable thing for a bad counterparty to acquire.
- **Report effectiveness, not activity**: clearance failure rate from the sample; lift over random, meaning how much better the alerts do than chance; and the fraction of your transaction space that has received zero non-random review, which is the leading indicator that illicit activity has moved into a segment your rules do not reach.

> **WHY THIS MATTERS**
>
> Every item on this list is measurable with data the program already holds. What is missing is not capability. It is the decision to measure the outcome instead of the activity.

Why This Is Now a False Claims Act Question

Where a contractor certifies compliance with an export-control, supply-chain, or cybersecurity requirement as a condition of payment, that certification is discoverable, and the screening behind it becomes evidence. This is not a hypothetical exposure. It is where the enforcement has gone.

The Department of Justice's Civil Cyber-Fraud Initiative, running since 2021, uses the False Claims Act against contractors that misrepresent their cybersecurity posture. The object being certified is a program, not an outcome, which is the point: no breach, no exfiltration, and no attack is required. The false statement is the attestation itself. In fiscal year 2025 the Department reported more than fifty-two million dollars recovered across nine cybersecurity settlements, part of a record year for whistleblower filings.[1]

The worked cases show the pattern precisely.

A logistics contractor self-reported a perfect cybersecurity score of 110. A later government assessment, run by the Defense Contract Management Agency's assessment center, scored the same company at negative 170 — a 280-point gap on a single self-attested number, against a possible range that bottoms out at negative 203. The company resolved the allegations for $507,144. No breach was alleged; the covered conduct was billing while knowing the required control set was not implemented.[2]

Another contractor self-reported 104, near the top of the range. An outside assessment told it the real figure was negative 142, reflecting roughly a fifth of the required controls in place. The company did not correct the reported score until months later, after a federal subpoena. It resolved the matter for $4.6 million.[3]

A research university's reported score was premised on an assessment environment that did not correspond to any actual system handling the government's information. It settled for $875,000. A second university settled for $1.25 million over controls it had represented it would remediate on a schedule it did not keep.[4]

These are settlements, and the allegations were resolved without any admission or adjudication of liability. But the pattern is unambiguous. A self-attested score, unaccompanied by any measurement of whether the attested controls actually work, is the cheapest-to-fake attribute in the whole system — and it is now the precise object around which enforcement is built.

> **WHY THIS MATTERS**
>
> When the government pays for a compliance program rather than a compliance outcome, the question of whether a misrepresentation was material turns on whether the program the contractor ran was the program the government paid for. A score with nothing behind it documents a program without evidencing one. The distinction is exactly what the random-sample audit is built to establish.

Measurement Is the Defensible Posture. Non-Measurement Is Not.

There is a tempting misreading of everything above, and it needs to be met directly, because it is the reading a hostile party would prefer you adopt. The misreading runs: if measuring my own failure rate creates a record of what I knew, it is safer not to measure. Under the False Claims Act's scienter standard — which reaches actual knowledge, deliberate ignorance, and reckless disregard[5] — that reasoning is not just wrong, it is dangerous.

Three reasons it is wrong.

First, the exposure that has actually cost companies money is the false certification, not the discovery of a residual risk. A program that measures its failure rate and acts on it — tightening screening where the sample shows leakage — converts the measurement into evidence of diligence. The adverse posture attaches to measuring and then doing nothing, which is a governance failure independent of any statute.

Second, materiality is a partly separate question, and disclosure builds a defense on it. Where the government continued to pay with knowledge of a disclosed weakness, that continued payment is evidence the requirement was not material — a defense at least one contractor has already argued in live litigation, on the theory that the agency never treated the certification as the essence of the bargain and kept paying after learning of the alleged noncompliance. That defense is available only to a contractor who measured and disclosed. Silence forfeits it.

Third, the alternative to measurement is not safety. It is a certification with no evidentiary basis, which collapses the moment the underlying conduct surfaces, with nothing to fall back on. Choosing not to look does not make the exposure smaller. It makes it undocumented, and a documented decision not to examine a known blind spot is closer to the deliberate-ignorance the statute reaches than to a safe harbor.

> **WHY THIS MATTERS**
>
> The instinct to avoid measurement in order to avoid knowledge produces the worst available position: the exposure remains, the diligence defense is gone, and the materiality defense is gone with it. Measurement is the posture that can be defended. Non-measurement is the one that cannot.

Three Questions That Settle It for Your Program

None of these requires counsel to answer, though the answers may send you to counsel.

One. Can we state our clearance failure rate? Not our alert volume or our clearance speed — the estimated rate of disqualifying conduct among the counterparties we cleared. If the honest answer is that there is no way to know, the program measures activity, not effectiveness, and it cannot substantiate a reasonable-care representation on the merits.

Two. Would our screening survive a counterparty that changed its name last quarter? If a re-registered entity clears as new, the program is not resolving entities to a persistent identity, and its record against the counterparties most likely to be concealing something is unknown to us.

Three. What sits behind our own — or our supplier's — attested compliance score? If the answer is the score itself and nothing that measures whether the controls actually work, then the most enforcement-exposed attribute in the whole program is also its least substantiated.

If those three answers are clear and consistent, the program can defend itself, and that is a real outcome — more common than the enforcement record suggests, because settlements are the only cases anyone publishes. If any one of them conflicts — a certification with nothing behind it, a screen that clears renamed shells, a program that cannot state what it missed — that conflict is the finding. It is far better identified before an examiner, an opposing expert, or a whistleblower identifies it.

When This Stops Being a Design Question

Whether a particular certification was accurate when it was made, whether a posted score can be substantiated, and what to do about a representation already submitted that may have been wrong are legal determinations, and the last of them belongs with counsel immediately. What is a design question, and what this assessment addresses, is whether a screening program is built to detect the conduct it exists to catch, and how to know. That question has an answer, and the answer is to measure the outcome rather than the activity.

This assessment is not legal advice and is not a substitute for counsel.

LIMITATIONS

What this assessment does not establish

This assessment describes a structural pattern in how screening programs are validated and a design change that corrects it. It does not determine whether any particular program is deficient, whether any particular certification was false, or how any specific matter would be decided. Those depend on facts and instruments not in the public record.

The random-sample audit estimates the failure rate against conduct that enhanced review can detect. It does not bound exposure to conduct that no available review method would catch, and it should not be represented as doing so.

The False Claims Act analysis describes a strategic structure under current authority. It is not a prediction of outcome in any matter, and the materiality defense described is probative but not dispositive: a government may continue to pay for reasons that do not render a misrepresentation immaterial, and courts have declined to resolve the weight of continued payment at the pleading stage.

The named enforcement matters are settlements. They are cited only for what their public settlement records establish, and the allegations were resolved without adjudication of liability. Settlements over-represent the resolutions the government chose to publicize and under-represent declinations and matters resolved under seal.

An assessment that cannot state what would change it should not be published. This one would change if a court held that process-metric validation satisfies a reasonable-care standard, if a Supreme Court decision altered the scienter or materiality tests the analysis relies on, or if the Civil Cyber-Fraud Initiative enforcement pattern reversed.

The full framework, with its statistical treatment, its entity-resolution detail, and its complete citations, is published as a companion academic paper on SSRN: *Adversarial-Cost Screening: A Validation Framework for Export Control, Sanctions, and Supply-Chain Counterparty Programs*. [SSRN URL — insert on approval of Abstract ID 7252300.]

SOURCES

Primary sources

Every factual and legal assertion traces to a primary statutory, judicial, or Department of Justice record. Where the public record does not settle a question, the text says so rather than resolving it by inference. Full citations, including the supporting scholarship, appear in the companion SSRN paper.

1. U.S. Department of Justice, *Fact Sheet: False Claims Act Settlements and Judgments, Fiscal Year 2025* (16 January 2026) (over $52 million recovered across nine cybersecurity fraud settlements; civil cyber settlements more than tripled in each of the past two years). **https://www.justice.gov/opa/media/1424126/dl**

2. Settlement agreement, *United States v. LOGZONE, Inc.* (N.D. Ala., June 2026), recording the October 2021 self-assessed score of 110 and the February 2024 DIBCAC assessment of negative 170; U.S. Department of Justice release of 18 June 2026. **https://www.justice.gov/opa/media/1446716/dl**

3. Settlement agreement, *United States ex rel. Berich v. MORSECORP, Inc.*, No. 23-cv-10130 (D. Mass.); U.S. Department of Justice release of 26 March 2025 (self-reported score of 104; third-party assessment of negative 142; correction delayed until after a federal subpoena; $4.6 million).

4. U.S. Department of Justice, settlement with Georgia Tech Research Corporation (30 September 2025, $875,000); settlement with the Pennsylvania State University (22 October 2024, $1.25 million).

5. *Universal Health Services, Inc. v. United States ex rel. Escobar*, 579 U.S. 176 (2016) (materiality is a demanding standard; continued government payment despite knowledge of noncompliance is strong evidence of immateriality); *United States ex rel. Schutte v. SuperValu Inc.*, 598 U.S. 739 (2023) (scienter turns on the defendant's subjective knowledge, and reaches actual knowledge, deliberate ignorance, and reckless disregard); *United States ex rel. Cimino v. IBM Corp.*, 3 F.4th 412 (D.C. Cir. 2021) (continued payment probative of immateriality but not dispositive at the pleading stage). False Claims Act, 31 U.S.C. §§ 3729–3733; scienter defined at § 3729(b)(1).

SUGGESTED CITATION

George, Collin B. *Cleared Does Not Mean Examined: Why a Compliance Screening Program Can Pass Every Audit and Still Not Detect the Thing It Exists to Catch.* Assessment SB-2026-03, version 1.0. Sanctir LLC, 9 August 2026.

METHOD

Method note

This assessment is independent open-source analysis. It applies two established results — selection on the dependent variable, from program evaluation, and strategic classification, from the machine-learning literature — to the validation of compliance screening programs, a domain in which neither is routinely applied. Its legal findings trace to primary sources: the False Claims Act at 31 U.S.C. 3729–3733; the Supreme Court and Circuit authority cited; and Department of Justice settlement records and fiscal-year statistics. Where the public record does not settle a question, that is stated rather than resolved by inference.

The full methodological treatment, including the statistical and entity-resolution detail omitted here for brevity, is published as a companion academic paper on SSRN. Sanctir is a solo practice; this work was subjected to adversarial self-review, not external peer review.

AUTHOR

About the author

Collin B. George, CISSP, is the principal of Sanctir LLC, an independent research and advisory practice working on CMMC and NIST SP 800-171, export controls, sanctions, and defense industrial base risk.

Sanctir is a solo practice. This assessment was subjected to adversarial self-review rather than external peer review, and is not affiliated with any government agency, academic institution, or defense contractor.

ORCID 0009-0007-8162-6839 · Full background · Contact