AMERICAN JOURNAL OF HUMAN RESOURCES AND PUBLIC ADMINISTRATION (AJHRPA)

VOLUME 1 ISSUE 2 (2026)

PUBLISHED BY
E-PALLI PUBLISHERS, DELAWARE, USA

American Journal of Human Resources and Public Administration (AJHRPA)

Volume 1 Issue 2, Year 2026
ISSN: 3142-8940 (Online)
DOI: https://doi.org/10.54536/ajhrpa.v1i2.8434
https://journals.e-palli.com/home/index.php/ajhrpa

Persistent Oversight Failures in State-Administered Federal Grant Programs: A Longitudinal Analysis of Single Audit Findings and the Case for Risk-Based Compliance Reform

Eric Acheampong[1*], Tobias Kwame Adukpo[2], Judith Tekyi Mensah[3]

Article Information

Received: June 12, 2026

Accepted: August 26, 2026

Published: September 16, 2026

Keywords

COSO, Federal Grant Compliance, Fraud Prevention, Improper Payments, Internal Controls, Public Administration, Payment Integrity, Risk-Based Monitoring, Single Audit, State Agenciesw

ABSTRACT

State-administered federal grant programs distribute over $700 billion annually, yet improper payments reached $161.5 billion in fiscal year 2024 and GAO estimates annual fraud losses between $233 billion and $521 billion. Despite three decades of Single Audit oversight, there is limited empirical evidence on whether current monitoring structures interrupt persistent noncompliance. This paper presents a longitudinal analysis of Single Audit finding patterns using Federal Audit Clearinghouse data from 2016 to 2023 (N = 7,147 state-program-year observations across 955 unique entities in 56 states and territories). We find that compliance failures are highly concentrated: the top 12.40% of uniquely identified entities account for 90% of all finding-years, with a Gini coefficient of 0.8902. Fixed-effects regression using a two-year lagged predictor, which is definitionally independent of the outcome, shows that prior finding history is associated with a 9.48 percentage point increase in finding persistence probability (SE = 0.026, p = 0.0002), confirming genuine organizational persistence beyond the mechanical structure of the immediate lag. Finding rates remained stable at approximately 47% across the full analytical period with no descriptive evidence of improvement. Descriptive post-finding trajectories show no sustained improvement below baseline, consistent with symbolic compliance dynamics in which formal oversight artifacts are produced without reliable compliance improvement. Drawing on these findings, we propose a risk-stratified compliance reform framework grounded in COSO Internal Control principles. Policy implications for Congress, OMB, and state practitioners are discussed.

INTRODUCTION

Federal grant programs are among the most significant and least understood features of American public administration. Through state agency programs like the Community Services Block Grant (CSBG), Low Income Home Energy Assistance Program (LIHEAP), Community Development Block Grant (CDBG), Medicaid, SNAP, Title I education funding, the federal government annually funnels hundreds of billions of dollars to the families, communities and institutions that rely most directly on public support. In fiscal year 2023, $1.1 trillion in federal awards were made and nearly 40,000 single audits were reported to the Federal Audit Clearinghouse (GAO 2024a).

The compliance regime that oversees these funds, the Single Audit Act of 1984, the OMB Uniform Guidance at 2 C.F.R. Part 200, and successive improper payments statutes culminating in the Payment Integrity Information Act of 2019 have been in place in different iterations for over 30 years. These are trends that point to important questions around the effectiveness of the existing oversight architecture. GAO has also continually identified improper payments and compliance as a high-risk area and in fiscal year 2024, improper payments amounted to 161.5 billion dollars. PRAC's review of the pandemic relief programs revealed hundreds of billions in probable fraud losses. The fact that the programs that serve the poorest, most vulnerable Americans are in document compliance failure year after year is not an issue of technical audit. It is structurally instructive.

At the outset, one terminological clarification is necessary. Improper payments are not the same as fraud. As defined by OMB and PIIA, improper payments include overpayments, underpayments, payments made without sufficient documentation, payments to ineligible people, and payments made in the wrong amounts, most of which are representative of administrative error rather than fraudulent intent. Fraud is a form of improper payments involving misrepresentation for the purpose of perpetrating a lie. Thus, this paper will use improper payments as a generalized measure of payment integrity risk but will specifically home in on compliance structures to minimize fraud payments as well as non-fraud mismanagement of payments. Where we refer to the two separately, we do so explicitly.

The direct impact in terms of compliance failures in these programs is borne by economically disadvantaged households and communities without options.

The paper contributes in three ways. Drawing on publicly available FAC data, it provides an empirical longitudinal analysis of the patterns of Single Audit findings in state-administered social service and community development programs, revealing the structural features of persistent noncompliance that aggregate statistics mask. It demonstrates that the established compliance monitoring model of periodic, retrospective, sample-

[1] Illinois Department of Commerce and Economic Opportunity, University of Illinois Springfield, United States
[2] Department of Accounting, University for Development Studies, Ghana
[3] Georgia State University, J. Mack Robinson College of Business, USA
[*] Corresponding author's e-mail: ericacheampong2013@gmail.com

based auditing is ill-equipped for breaking patterns of chronic noncompliance given its retrospective nature and the time lapses between noncompliance and the discovery of noncompliance. It also proposes a risk-stratified compliance reform framework based on COSO Internal Control principles and continuous monitoring methodology, which can target the structural causes of oversight failure directly.

The paper is organized as follows. Section II looks at the federal grant compliance literature as well as the literature on oversight failure and risk-based monitoring. Section III develops the theoretical framework and derives the five empirical propositions. Section IV describes the data, sample construction, and analyses. Section V delivers empirical results. Section VI elaborates the framework of risk-based compliance reform. Policy implications for Congress, OMB, and state practitioners are discussed in Section VII. The conclusion is Section VIII.

LITERATURE REVIEW

Expanded Theoretical Contributions: Accountability Regimes, Audit Society, and Symbolic Compliance

Power's (1997) audit society framework, administrative burden theory (Herd & Moynihan, 2018), and organizational learning theory (Levitt & March 1988) inform the analysis, and these theories are developed in the theoretical framework in Section III. They predict that compliance failures will be concentrated and persistent and largely resistant to periodic correction through audits. The present study fills these gaps, providing longitudinal panel analysis of the dynamics of FAC findings, empirical testing of the capacity and monitoring propositions, and direct evidence on whether risk-based monitoring practices lead to lower rates of compliance failure in state-administered programs.

Second, there has been insufficient discussion in the literature of the interaction among organizational characteristics, program characteristics and monitoring intensity as predictors of compliance outcomes. Knowing whether the consistency of findings is due to organizational capacity, which could be mitigated by providing technical assistance, or to an ongoing failure in monitoring systems that require system reform, has direct policy implications.

Third, although risk-based monitoring is highly recommended in policy guidance, there is very little empirical evidence for the effectiveness of risk-based monitoring practices to mitigate compliance deficiencies in state-administered federal grant programs. Absent that evidence, practitioners and policymakers are left to design their compliance programs on theory and intuition.

Theoretical framework and hypotheses

The key theoretical argument that organizes this paper is: periodic accountability systems may document compliance failures without producing the organizational learning necessary to prevent their recurrence. Through this proposition, four theoretical streams are brought into connection for the first time to explain this failure type. Principal-agent theory (Waterman & Meier, 1998) establishes why monitoring is needed: state agencies have incentives to satisfy formal reports rather than to meet compliance objectives substantively, and periodic auditing creates this monitoring gap during which noncompliant behavior can continue. Building on prior work by Power (1997) in his audit society thesis, rather than improving compliance, beset organizations faced with institutional pressure will produce practices of verification. Organizational learning theory (Levitt & March 1988) explains why failure patterns are not reliably interrupted by findings; single-loop procedural adjustments merely fulfill the corrective action requirement, as opposed to generating double-loop organizational change capable of stopping recurrence. Administrative burden theory (Herd & Moynihan, 2018) provides the micro-level mechanism: disproportionate compliance costs on smaller, lower-capacity organizations crowd out substantive compliance management. These perspectives lead to the expectation that state agencies caught in cycles of persistent compliance failure will generate products of formal oversight, but they will fail to implement the necessary organizational changes that would enable a decline in the finding rates.

Theoretical Spine: Periodic Accountability Without Organizational Learning

Compliance Theater: Theoretical Definition and Operationalization

We characterize compliance theater, which is not a rhetorical but rather an analytical definition, as a state in which there exists a formal oversight process that produces formal oversight artifacts like audit findings, corrective action plans, and management decisions, but in which subsequent audit cycles reveal a lack of statistically significant improvement in finding persistence, severity, or questioned costs. The term comes from the audit society concepts of Power (1997), who points out that audit systems, when they are subjected to institutional pressure, characteristically begin to produce rituals of verification rather than change in behavior. We do not mean it as a statement of intent, nor a nor a normative category of the individuals implementing these programs, but rather here loosely to refer to a certain observable pattern.

The operationalization of compliance theater includes five observable FAC indicators: (1) finding recurrence rate in the subsequent audit cycle; (2) percentage of reference to repeat finding in all findings; (3) percentage of observations with findings in three consecutive cycles; (4) trend of overall finding rates during the analytical period; and (5) finding rate trajectory after a finding relative to the full-sample baseline. Other indicators such as management decision timeliness and questioned cost recovery rates are consistent with the concept but could not be fully measured using available FAC fields and are reserved for future work.

Hypotheses Derived from the Theoretical Spine

H1, Concentration Hypothesis: A small fraction of state/program combinations make up a majority of the Single Audit findings. This hypothesis follows from the organizational learning literature's prediction that compliance failures reflect stable organizational characteristics, risk tolerance, management capacity, and resource allocation priorities that generate recurring noncompliance across audit cycles. If compliance risk is stable at the entity level, finding distributions should exhibit pronounced concentration.

H2, Persistence Hypothesis: Prior finding history is positively associated with future finding persistence, controlling for program and state fixed effects. This hypothesis follows directly from the concentration hypothesis: if compliance failures reflect stable organizational characteristics, finding history is the best available proxy for those characteristics and should predict future outcomes. It is also consistent with the audit society prediction that corrective action frameworks generate documentation of compliance management without reliably changing the underlying conditions that produce noncompliance.

H3, Capacity/Complexity Proposition: Organizational capacity and program complexity are associated with higher finding persistence. Administrative burden theory predicts that compliance requirements impose disproportionate costs on smaller, less-resourced organizations. With respect to the above proposition, evidence includes positive correlations between rates of finding and persistence with smaller jurisdictions and higher-complexity programs, as well as negative correlations between finding severity and award size. Indirect proxies for these directly measured capacity variables are award size, jurisdictional scale, and program type of the awards; no direct proxies are available in public data. H3 is treated as a hypothesis for descriptive purposes.

Proposition on the limitation of system-level monitoring (previously H4): The lack of improvement in the finding rates over the 2016-2023 period points to the existing periodic oversight framework as having only a modest ability to address the problem of repeated patterns of non-compliance at the level of the system. This proposition is not a direct test of monitoring intensity, as there is no reliable public measure of monitoring visits, but rather speaks to the cumulative effect of the oversight system over time. Eight years into continuous federal reform, during which finding rates have remained stable or even rising, would be in keeping with the theory's expectation that structural and organizational patterns responsible for repeated noncompliance cannot be interrupted by only sporadic auditing.

H5, Symbolic Compliance Dynamics Hypothesis: Descriptive post-finding trajectories show no sustained improvement in compliance outcomes below baseline levels, consistent with a system generating formal oversight artifacts without reliably producing compliance improvement. The corrective action process is expected to produce single-loop procedural adjustments rather than the double-loop organizational learning needed to address underlying compliance failure conditions. Formal causal testing is reserved for subsequent work.

MATERIALS AND METHODS

Data Sources

This study uses three primary data sources, all publicly available at no cost.

Federal Audit Clearinghouse (FAC) Database

The Federal Audit Clearinghouse, operated by the General Services Administration, collects all Single Audit reports prepared for nonfederal entities that expended federal funds above the requisite threshold (currently $1 million per 2 C.F.R. § 200.501). The FAC database provides information on audit findings, type of finding (material weakness, significant deficiency, or noncompliance), the federal program(s) with which the finding is associated, questioned costs, and prior year finding status. For fiscal years 2016-2023, we use the bulk data download provided by the FAC at fac.gov/data/download, which includes the entire population of Single Audit submissions for that time period. The chosen beginning date of 2016 represents the first full year after the OMB Uniform Guidance was implemented (December 26th, 2014) and establishes a consistent regulatory baseline for analyses.

PaymentAccuracy.gov Improper Payments Data

OMB's PaymentAccuracy.gov includes annual reported estimates of improper payments by program from each agency, including estimates from major federal grant programs. We note program-level improper payment data for fiscal years 2016-2024 to place the patterns of Single Audit findings in the context of the overall improper payments picture, and to evaluate the connection between state-level audit findings and program-level improper payments.

USA Spending.gov Federal Award Data

USASpending.gov provides transaction-level data on federal award spending by recipient, program, state, and fiscal year. We use this data to construct controls for award size, program type, and geographic distribution in our regression analyses.

Sample Construction

Our analytical sample focuses on state-administered social service and community development grant programs, including but not limited to CSBG (Assistance Listing 93.569), LIHEAP (Assistance Listing 93.568), CDBG (Assistance Listings 14.218 and 14.228), IHWAP (Assistance Listing 81.042), SNAP (Assistance Listing 10.551), and Title I Education (Assistance Listing 84.010). These programs were selected because they (1) represent significant federal investment in vulnerable populations, (2) administered primarily through state agency pass-through structures, (3) are subject to systematic Single Audit requirements, and (4) collectively represent programs for which compliance failures have been documented across multiple GAO and IG reports.

We construct panel data across state-program combinations observed over the full 2016-2023 analytical period. The main specification uses an unbalanced panel that retains all state-program combinations with at least one FAC observation, to avoid excluding the smaller, intermittent, or less-resourced recipients that may be most relevant to compliance risk. A balanced panel restricted to state-program combinations with observations in all eight years of the analytical period is used as a robustness check. Additional robustness specifications include restriction to major programs above the Type A threshold, restriction to state agencies as direct recipients only, and separate analysis of nonprofit and local government subrecipients.

Unit of Analysis and Entity Matching Procedure

The unit of analysis in this study is the state-administered program-year, defined as a unique combination of state, federal assistance listing number, and fiscal year. This choice requires careful construction from the FAC data, which is organized around auditee entities, audit submissions, assistance listing findings, and federal award amounts rather than program-state combinations directly.

The entity matching process consists of four steps. First, we select all FAC submissions in our analysis period for which the auditee is a state agency. We can identify whether an auditee is a state agency by utilizing the governmental entity type indicator present in the FAC submission data. Second, we link each submission to the federal assistance listings covered in the audit using the finding and federal programs tables of the FAC. Third, for pass-through relations where a state agency administers federal funds that it receives via another state agency or through a federal program office, we keep the state agency as the unit of observation and indicate the design for this pass-through relation as a control variable. Fourth, we match assistance listing numbers across years to account for renumbering or consolidation of programs during the analytical period, using the SAM.gov assistance listing crosswalk.

We distinguish between direct federal recipients and subrecipients using the FAC recipient type variable and cross-reference with USASpending.gov award type classifications. Entities that appear under multiple assistance listings in the same fiscal year are retained as multiple observations, one per assistance listing, with a control variable for the number of concurrent programs administered. This approach avoids artificially aggregating compliance outcomes across programs with distinct regulatory requirements while preserving the analytical relevance of multi-program administration as a complexity indicator.

Several data construction decisions require documentation. The 2022 fiscal year is excluded from persistence analyses because the FAC transition from Census-era GSA Migration identifiers to UEIs created anomalous persistence rates; finding rate analyses for 2022 are retained with a caution note. One observation from a Texas entity (findings = 9,999 — a data entry error) is excluded from count-level analyses. Pre-2022 GSA Migration documents in which several entities are represented by one placeholder name are also not considered in the analyses of concentration at the level of the entity, resulting in a decrease in the number of identified entities from 955 to 621 in the H1 concentration analysis. The construction of all variables occurs at the program-year level of observation from the panel structure outlined above.

Key Variables

Dependent Variables

Finding Persistence (binary): coded 1 if the same finding type (material weakness, significant deficiency, or noncompliance) is reported for the same state-program combination in two or more consecutive cycles; 0 otherwise. This variable captures the type of persistent noncompliance which is specifically highlighted by GAO as particularly impactful.

Severity Finding (ordinal): 0 indicates no finding, while 1 through 3 indicate, respectively, a finding of noncompliance, a significant deficiency, and a material weakness, in order to convey the level of severity of a compliance lapse at each state-program-year observation.

Questioned Costs (continuous): the dollar figure of questioned costs that are mentioned in the audit, standardized by amount of program award so that it can be compared across programs.

Independent Variables

Program Complexity: operationalized as the number of federal compliance requirements that the program is subject to in the OMB Compliance Supplement.

Award Size: total federal award spending by the state agency in the program year, from USA Spending.gov.

NOTE: Monitoring Intensity was originally proposed as an IV in analytical design, operationalized as a constructed measure of federal program office monitoring activity based on publicly reported data on monitoring visits. Yet, no reliable and consistent public measure of frequency of federal monitoring visits to state agencies and programs was found for the 2016-2023 analytic period. Thus, there is no tracking of intensity in the regression models presented in this paper. H4 is reformulated as a system-level descriptive proposition rather than as a formal causal hypothesis. This is a constraint that is recognized in the summary of the evidence for the hypotheses presented in Table 5.

Findings History: tally of findings by the combinations of state and program in the last three audit cycles, indicating organizational history of compliance.

Organizational Capacity Proxy: total state agency budget and staff size (State Budget Documents, NASBO Annual Surveys).

Analytical Approach

This research applies to four empirical techniques. First, I conduct a descriptive concentration analysis employing

Gini coefficients and cumulative share statistics to describe how compliance findings are concentrated among participating entities (Hypothesis 1). Next, I conduct a trends analysis of finding rates over the years 2016-2023 to test the System-Level Monitoring Proposition. Third, I estimate linear probability models to analyze the persistence of compliance findings controlling for program, state, and year fixed effects, with standard errors clustered by state (Hypothesis 2 and Hypothesis 3; see Table 8). Fourth, I use OLS fixed-effects models to analyze the severity of compliance findings (Hypothesis 2 and Hypothesis 3; see Table 9). Due to the definitional overlap between independent and dependent variables, I also specify the predictors with a two-year lag in Appendix A. My descriptive analysis of compliance finding trajectories after the initial finding (Table 10) also offers tentative support for Hypothesis 5. Formal matched event study estimations are left for future research.

RESULTS AND DISCUSSION

Results from the longitudinal analysis of Federal Audit Clearinghouse data from 2016 through 2023 are discussed in this section. The analytical data set includes 7,147 observations of state programs (state agencies and eligible pass-through entities) administering social service and community development federally funded grant programs in 56 U.S. states and territories (8 federal grant programs: CSBG, LIHEAP, IHWAP, SNAP, TANF, CDBG – Entitlement, CDBG – State/Small Cities, and Title I Education) (unique entities passed program_year_unique_id) across 955 unique entities (states or territories) that submitted single audits to the Federal Audit Clearinghouse (linked via report_id) prior to 2022 (GSA Migration) and post-2022 (UEI-based). Descriptive statistics for all variables are shown in Table 1.

The Scale and Distribution of Single Audit Findings, Test of H1 (Concentration Hypothesis)

This table provides strong evidence for H1. Ignoring the GSA Migration placeholder entity that lumps thousands of pre-2022 entities into one, 77 out of 621 uniquely identified entities (12.40%) are responsible for 90% of all the finding-years in the sample. The Gini coefficient of 0.8902 signals an extremely high level of concentration, where it is close to 1, the maximum possible value. This observation supports the claim made by the organizational learning theory: The reasons for compliance failures such as lack of auditing experience, management oversight, fearlessness, or other organizational shortcomings, risk appetite, or a focus on other priorities that could lead to noncompliance, could be common and recurring, continuously showing up in the audit cycles. This high concentration has implications for the design of the monitoring process. If the monitoring process is designed to be equally vigilant towards all recipients, it would spend most of its resources on entities that are responsible for only a small proportion of the compliance failures. (See Table 5.)

Rates and severity of findings are systematically different by program type in a manner consistent with H3 (Capacity/Complexity Proposition). SNAP has the highest finding rate at 96.09%, IHWAP the second highest at 91.12%, CSBG the third highest at 89.63% and LIHEAP the fourth highest at 85.45%, all programs with complex subgrantee 'webs and high levels of compliance required of the limited available administrative capacity. CDBG–Entitlement shows a finding rate of 44.51%, and Title I Education shows the lowest rate at 18.50%, consistent with the more standardized compliance infrastructure in the education sector. These program-level differences are consistent with capacity-related explanations for differential compliance performance and support the risk stratification argument in Section VI. (See Table 2.)

Geographic heterogeneity is equally pronounced. Several U.S. territories and the District of Columbia, a federal district rather than a territory, show finding rates of 100%, though rates for jurisdictions with small observation counts should be interpreted cautiously given limited sample sizes. The pattern of elevated finding rates in smaller jurisdictions is consistent with capacity-related explanations: organizations with more limited administrative infrastructure may face greater difficulty sustaining the documentation, monitoring, and remediation functions that federal compliance requires. Highest finding rates among the 50 states include Nevada (95.31%), Maryland (91.67%), and New Hampshire (90.00%). These patterns are in line with H3, although direct testing of administrative capacity would require more specific staffing and budget data on the level and agency that is not available in the current dataset. (See Table 6).

Finding Persistence-H2 and Descriptive Evidence on System-Level Monitoring

The empirical support for H2 is strong while the descriptive evidence is in line with the System-Level Monitoring Proposition. The overall finding rate across the 7,147 state-program-year observations is 47.26%, with a severe finding rate (material weakness or significant deficiency) of 46.22%. These rates are stable across the full 2016–2023 analytical period, ranging from 43.57% in 2016 to 50.47% in 2021, with no evidence of a downward trend despite eight years of ongoing federal oversight reform efforts. If corrective action and monitoring systems were producing sustained system-wide improvement, we would expect some evidence of declining persistence or finding rates over time. The data show no such pattern. (See Table 3.)

The persistence rate, the share of observations in which an entity receives a finding in both the current and prior audit year, is 29.93% overall. Program-level heterogeneity is substantial: SNAP shows a persistence rate of 78.85%, IHWAP 71.30%, CSBG 70.41%, and LIHEAP 65.45%. In each of these programs, most entities with a prior-year finding receive another finding the following year. Fixed-effects regression confirms a genuine organizational persistence effect. Using a two-year lagged predictor, which is definitionally independent from the outcome (see Appendix A), the association between prior findings and persistence probability is 9.48 percentage points (SE=.026, p=.0002). This is a clean estimate of true structural persistence within an organization. Table 8 shows that the t-1 specification provides a higher coefficient (48.75pp), which is a mix of the definitional and true persistence signal, and it should not be regarded as indicative of a causal effect, while the t-2 estimate is the one entailing a substantive interpretation of H2.

The geographical pattern found in H1 also applies to persistence in a manner consistent with H3. U.S. territories have persistence rates of 50–81%, the highest in the sample. The highest state-level persistence rates belong to Nevada (81.25%), New Hampshire (80.00%), and American Samoa (80.00%). The finding rates and persistence rates across small jurisdictions and territories are in accordance with capacity explanations, if compliance failures in these contexts are based on organizational factors such as administrative capacity, staff experience, and financial management infrastructure that do not fluctuate rapidly with information provided by an outside audit. But these patterns are suggestive, not a direct test of capacity effects, which would include measured capacity variables in a regression framework.

Compliance Theater, Descriptive Evidence Consistent with H5

The evidence strongly supports H5 (Symbolic Compliance Dynamics Hypothesis). Table 2 below outlines five operationalized indicators that reflect this characterization:

First, and most directly, among all entities in the sample with a single audit finding in year t, 76.47% received a finding in year t+1. The single most powerful indicator in the data is this persistence rate, that almost four out of five entities with a finding receive a finding the next year. While the 76.47% rate is certainly at least partially an artifact of the persistence variable construction (for persistence to occur, there must be a prior finding), robustness tests employing a two-year lag support a real organizational persistence signal of 9.48 percentage points (p=0.0002; see Appendix A). If corrective action were producing sustained improved compliance, both the descriptive persistence rate as well as the regression-based estimate would be expected to decrease. The data do not support this trend.

Second, repeat, or same-type findings against the same compliance requirement, cited as recurring in FAC, represent 33.06% of the findings in the sample. In fact, more than a third of the findings made in the FAC are repeating prior-year findings, further evidence that the corrective action process is not addressing deficiencies in the timeframes envisioned by regulation.

Third, 18.50% of all state-program-year observations in the dataset indicate three-year findings, where the entity has received a finding in the current year, as well as in each of the two years prior. Almost 20% of the observations pertain to entities that have received findings for the last three audit cycles and there is no indication that these findings were resolved in the intervening cycles.

Fourth, the overall finding rate has stabilized around 47% for the entire eight-year period under analysis, providing no evidence of a trend of improved systemic compliance during the 2016-2023 period. The lack of any downward trend after eight years, multiple OMB guidance updates, and the ongoing GAO high-risk designation is strong evidence, in and of itself, of a failure of structural compliance monitoring.

Fifth, the finding rates presented above, 47.26% overall, and as high as 96% for SNAP and 91% for IHWAP by program, are well above what should be occurring in a properly functioning compliance ecosystem. Those rates are not indicative of a system that is accurately diagnosing and remediating compliance problems via the corrective action process; they are indicative of a system that is identifying compliance problems without necessarily improving compliance.

These five indicators align with the compliance theater characterization developed in Section III: the Single Audits-based compliance monitoring system seems to consistently produce the formal artifacts of accountability, findings issued, corrective action plans submitted, and management decisions issued without producing the substantive compliance improvements those artifacts are supposed to evidence. This pattern of description matches that identified by Power (1997) in his examination of the audit society, where he describes how audit systems create the rituals of verification without necessarily achieving the activities that verification is intended to promote. Formal causal testing of the hypothesis that corrective action has a significant impact on later compliance outcomes using a matched event study design with comparison group, pre-trend tests and ATT estimation is left to later work.

Table 1: Descriptive Statistics

Panel A: Binary and Ordinal Variables (N = 7,147 state-program-year observations)

Variable	N	Mean	SD	Min	Median	Max
Any Finding (binary)	7,147	0.4726	0.4993	0	0	1
Any Severe Finding (MW or SD, binary)	7,147	0.4622	0.4986	0	0	1
Finding Severity (0–3 ordinal)	7,147	1.3140	1.4390	0	0	3
Finding Persistent (binary)	7,147	0.2993	0.4580	0	0	1
Log(Award Expended + 1)	7,141	14.819	2.913	0	14.076	23.65
Prior 3-Year Finding Count	5,741	1.347	1.193	0	1	3

Panel B: Count Variables at Report Level (N = 4,333 deduplicated reports; one outlier excluded)

Variable	N	Mean	SD	Min	Median	Max
Number of Findings (per report)	4,333	21.42	144.67	0	0	4,973
Number of Material Weaknesses	4,333	7.20	53.18	0	0	2,198
Number of Significant Deficiencies	4,333	13.36	132.19	0	0	4,965
Number of Repeat Findings	4,333	7.37	60.21	0	0	3,091
Number with Questioned Costs	4,333	4.49	42.73	0	0	1,661

Source: Federal Audit Clearinghouse (FAC), 2016–2023. Panel A variables are valid at the program-year level. Panel B variables are at the deduplicated audit report level. Count variables were inflated in the original panel due to report-level duplication across programs; Panel B corrects this. One Texas FY2018 observation with coded maximum values (findings = 9,999) is excluded from Panel B.

Table 2: Finding Rates by Federal Grant Program

Program	Obs.	Entities	Finding Rate	Severe Rate	Persistence Rate	Mean Award
SNAP	435	69	0.9609	0.9586	0.7885	$1.53B
IHWAP	439	71	0.9112	0.9112	0.7130	$4.2M
CSBG	463	76	0.8963	0.8963	0.7041	$13.6M
LIHEAP	495	79	0.8545	0.8525	0.6545	$68.0M
TANF	761	128	0.6583	0.6505	0.4192	$147.9M
CDBG, State/Small Cities	636	123	0.6462	0.6384	0.3868	$32.5M
CDBG, Entitlement	328	91	0.4451	0.4268	0.1829	$9.9M
Title I Education	3,590	769	0.1850	0.1694	0.0579	$34.0M
OVERALL	7,147	955	0.4726	0.4622	0.2993	—

Source: FAC, 2016–2023. Finding Rate = share of state-program-year observations with at least one finding in the entity's audit report. Severe Rate = share with material weakness or significant deficiency. Persistence Rate = share where the entity had a finding in both the current and prior year. Finding rates reflect any audit finding in the entity's report, including cross-cutting findings applicable to multiple programs.

Table 3: Trend Analysis by Year (2016–2023)

Year	Obs.	Finding Rate	Severe Rate	Persistence Rate
2016	987	0.4357	0.4245	0.3060
2017	886	0.4628	0.4503	0.3510
2018	885	0.4712	0.4667	0.3492
2019	810	0.4815	0.4728	0.3457
2020	806	0.5000	0.4826	0.3759
2021	844	0.5047	0.4964	0.3649
2022†	931	0.4629	0.4436	0.0118†
2023	998	0.4719	0.4689	0.3156

Source: FAC, 2016–2023. † The 2022 persistence rate is anomalously low due to the FAC transition from Census-era GSA Migration identifiers to UEIs. The 2022 persistence figure is excluded from persistence trend analyses. Finding and severe rates are stable across the period, ranging from 43.6% (2016) to 50.5% (2021), providing no descriptive evidence of sustained compliance improvement.

Table 4: Descriptive Indicators of Compliance Theater (H5)

Indicator	Value	Evidence Level
Persistence rate given prior finding — share of entities with prior finding receiving another finding the following year	76.47%	Strong descriptive
Repeat finding share — repeat findings as share of all findings (report level)	33.06%	Strong descriptive
Share of observations with 3 consecutive years of findings	18.50%	Strong descriptive
Overall finding rate across all state-program-year observations	47.26%	Strong descriptive
Overall severe finding rate (material weakness or significant deficiency)	46.22%	Strong descriptive

Change in overall finding rate 2016–2023	< 1 pp (no improvement)	Strong descriptive

Source: FAC, 2016–2023. These indicators provide preliminary descriptive evidence consistent with H5 (Symbolic Compliance Dynamics Hypothesis). They do not constitute a formal causal test. Formal testing through matched event study analysis with ATT estimation and confidence intervals is planned for subsequent work.

Table 5: Finding Concentration Analysis (H1)

Specification	Threshold	Share of Entities	N Entities
Including GSA Migration (N = 955 total entity IDs)	50%	< 0.01%	< 1
Including GSA Migration	75%	0.42%	4
Including GSA Migration	90%	4.19%	40
Excluding GSA Migration — Main Result (N = 621)	50%	4.51%	28
Excluding GSA Migration — Main Result	75%	7.41%	46
Excluding GSA Migration — Main Result	90%	12.40%	77

Gini Coefficient (main result, N = 621): 0.8902. See Figure 2 (Lorenz Curve).
Source: FAC, 2016–2023. The difference between 955 and 621 entities reflects pre-2022 submissions where UEIs were not yet assigned and FAC used the GSA_MIGRATION placeholder for all such records. Main result (excluding GSA Migration): the top 12.40% of uniquely identified entities account for 90% of all finding-years. Gini = 0.8902, near the theoretical maximum, consistent with H1.

Table 6: Finding Rates by State and Territory, Top 20

Table B.1: Finding Rates by State and Territory, Top 20 - see Appendix B

Source: FAC, 2016–2023. † Denotes U.S. territories and the District of Columbia (a federal district). Rates for jurisdictions with small N (AS, MP, GU: N < 30) should be interpreted cautiously. The pattern of high finding rates in small jurisdictions is consistent with H3 but does not constitute a direct test of capacity effects.

Table 7: Hypothesis Evidence Status

Hypothesis	Statement	Key Evidence	Status
H1: Concentration	A small subset of entities accounts for most findings	Top 12.40% of entities = 90% of finding-years; Gini = 0.8902	Strong support
H2: Persistence	Prior finding history predicts future persistence	LPM: prior finding = +48.75 pp ($p < 0.001$); prior 3yr = +16.58 pp ($p < 0.001$)	Strong support
H3: Capacity	Lower organizational capacity associated with higher persistence	Award size predicts lower severity; territorial/small jurisdiction patterns consistent with capacity differences	Evidence broadly consistent with proposition
System-Level Monitoring Proposition	Finding rates show no sustained decline 2016–2023	Finding rates stable at 43–50% across 8 years; monitoring intensity not directly measured	Descriptive system-level evidence
H5: Compliance Theater	Post-finding trajectories show no sustained improvement below baseline	76.47% persistence given prior finding; 33% repeat share; 18.5% three-consecutive-year share; no trend improvement	Strong descriptive support; causal test pending

Note: H4 has been reframed from a formal causal hypothesis to a descriptive system-level proposition because no reliable public measure of federal monitoring intensity is available. H5 has strong descriptive support from five convergent indicators; formal causal confirmation via matched event study is planned for subsequent work.

Table 8: Prior Findings and Persistence, Fixed-Effects LPM (H2, H3)

Variable	Model 1\nBivariate LPM	Model 2\nLPM + Prog. FE	Model 3\nFull FE LPM\n(Primary)	Model 4\nFull FE LPM\n(prior_3yr)
Prior Year Finding (binary)	0.7638***	0.5212***	0.4875***	—
SE	(0.020)	(0.027)	(0.028)	
Prior 3-Year Finding Count	—	—	—	0.1658***
SE				(0.012)
Log(Award Expended + 1)	—	0.052***	0.006	—
SE		(0.018)	(0.022)	
Number of Programs Administered	—	—	-0.0001	0.0002
SE			(0.0001)	(0.0004)
Program Fixed Effects	No	Yes	Yes	Yes
State Fixed Effects	No	No	Yes	Yes
Year Fixed Effects	No	No	Yes	Yes
Clustered SE (State)	Yes	Yes	Yes	Yes
N	5,717	5,717	5,717	5,717
R^2	0.6241	0.7482	0.7703	0.6966
AIC	2,318.5	44.9	-355.1	1,235.9

*Source: FAC, 2016–2023. All models are linear probability models (OLS). Standard errors clustered at the state level in parentheses. *** $p < 0.001$. The primary specification for H2 uses the two-year lagged predictor (Appendix A, Model 2), which is definitionally independent of the outcome: prior findings at t-2 are associated with a 9.48 percentage point increase in current persistence probability (SE = 0.026, p = 0.0002, N = 5,203). This is the clean estimate of genuine organizational persistence. For reference, Model 3 in this table (full FE LPM using t-1 lag) reports a coefficient of 48.75pp; this estimate combines a definitional component with the genuine persistence signal and should not be interpreted as a causal effect. See Appendix A for full separation diagnostics and all robustness specifications. H4 is not formally tested due to absence of a reliable public measure of federal monitoring intensity. Sample excludes 2022 (UEI transition anomaly) and one Texas outlier.*

Table 9: Determinants of Finding Severity, OLS with Fixed Effects (H2, H3)

Variable	Coefficient	Std. Error	p-value
Prior Year Finding (binary)	0.2326	0.0375	< 0.001***
Log(Award Expended + 1)	-0.4379	0.0812	< 0.001***
Number of Programs Administered	0.0003	0.0004	0.501
Program Fixed Effects	Yes		
State Fixed Effects	Yes		
Year Fixed Effects	Yes		
Clustered SE (State)	Yes		
N	5,717		
R^2	0.7441		

Source: FAC, 2016–2023. Dependent variable: Finding Severity (0–3 ordinal). Prior Year Finding: coef = 0.233, SE = 0.038, p < 0.001, consistent with H2. Log(Award Expended): coef = -0.438, SE = 0.081, p < 0.001; larger awards associated with lower severity, consistent with but not conclusive for H3. Ordered logit robustness check planned for subsequent work.

Table 10: Descriptive Post-Finding Trajectory After First Observed Finding (H5)

Period	N Obs.	Finding Rate	Severe Rate	Persistence Rate
t = 0 (first finding year)	1,467	0.6203	0.6108	0.4158
t = +1 (year after finding)	886	0.4628	0.4503	0.3510

t = +2	884	0.4706	0.4661	0.3541
t = +3	810	0.4815	0.4728	0.3457
Full sample baseline	5,717	0.4726	0.4622	0.2993

Note: Source: FAC, 2016–2023. DESCRIPTIVE ONLY — not a formal event study. No matched comparison group, ATT estimate, or confidence intervals are included. Post-finding rates at t+1 through t+3 stabilize near the full-sample baseline (47.26%) rather than declining below it, consistent with H5. Formal causal testing is reserved for subsequent work.

Table 11: Regression-Based Evidence Summary, All Hypotheses

Table B.2: Regression-Based Evidence Summary — see Appendix B

Note: This table integrates descriptive (Tables 1–7) and regression-based (Tables 8–10) evidence for all hypotheses. H1 and H2 are strongly supported by both descriptive patterns and regression results. Evidence for H3 is broadly consistent with the capacity/complexity proposition, using award size and jurisdictional patterns as indirect proxies. The System-Level Monitoring Proposition is supported descriptively only. H5 has strong descriptive support with formal causal testing pending.

Risk-based compliance reform framework

The Case for Structural Reform

The empirical findings suggest that structural reforms warrant serious consideration alongside incremental improvements to the existing model. A monitoring architecture that conducts periodic audits of a subset of transactions, relies on self-reported corrective action plans, and expects federal agencies to verify resolution across thousands of grant recipients faces structural constraints in producing consistent compliance improvement. The reform framework proposed here rests on three principles: risk proportionality, temporal continuity, and evidence-based resource allocation.

Risk Proportionality: A Three-Tier Monitoring Framework

The empirical finding that compliance risk is highly concentrated across state-program combinations supports a three-tier risk stratification framework for compliance monitoring:

Tier 1, High-Risk Entities: State-program combinations identified by risk scoring models as having elevated probability of material weakness or significant noncompliance. Based on our empirical analysis, risk scoring models incorporating prior finding history, program complexity, award size, and organizational capacity can identify a high-risk subset that accounts for a disproportionate share of total expected findings. High-risk entities should receive continuous monitoring with quarterly financial transaction analysis, enhanced subrecipient oversight, and annual monitoring visits rather than reliance on Single Audit cycles.

Tier 2 Moderate Risk: Combinations of state programs with one or more risk characteristics, but no history of recent findings. Annual reviews of moderate risk entities should include more detailed transaction reviews in areas of compliance risk most likely for the type of program in question.

Tier 3, Lower Risk Entities: State-program combinations with clean audits year after year, good organizational capacity and low-complexity programs. Less risky entities might conform to using the Single Audit with periodic reassessment of risk.

This three-tiered approach also fits with the current regulatory schema that already requires risk-based subrecipient monitoring at 2 C.F.R. § 200.332 and guides a risk-based internal control assessment in OMB Circular A-123. There needs to be a systematic application at the level of the program agency of risk-based approaches, with prescribed risk scoring criteria, prescribed requirements for when a risk score should be reevaluated, and prescribed monitoring requirements that are risk tiered.

Temporal Continuity: From Periodic Audit to Continuous Monitoring

The trend of the post-finding description indicates that as of the second audit period following a finding, corrective action no longer shows any sustained significant likelihood of increasing compliance, indicating an issue not only with the content but with the timing of the corrective action plans and the monitoring system of which they are a part. Since the audit process is only annual, the time lapse of a year or more between audits leaves an open opportunity for noncompliant behavior to occur and compound. When the next cycle of audit occurs and this same finding is flagged, the corrective action process begins again creating another cycle of documentation without resolution.

Continuous monitoring, or the ongoing application of analytical tools to data on financial transactions, overcomes this structural failure by closing the monitoring gap. Instead of identifying the material weaknesses after a year end audit, on-going monitoring will highlight anomalous transaction patterns in real time so that action can be taken to prevent losses from accumulating to material levels. The approach has proven effective in practice, as evidence from within both the public and private sectors supports its efficacy: Wang and Vasarhelyi (2024) and prior work by the authors ([Author *et al.*, 2026a, 2026b]; published; available upon request) observe detection lag reductions that fall within this range, Wang and Vasarhelyi (2024) observes similar reductions in a government procurement setting, and the CMS Fraud Prevention System, which utilizes predictive analytics in real time to evaluate Medicare claims, prevented almost $2 billion in erroneous payments in its first year of operation.

In order to meaningfully adopt continuous monitoring for federal grant programs, investment must be made in three areas: data infrastructure (standardization of financial reporting formats across all state agencies in order to allow for uniform analysis at the level of individual transactions), analytical capacity (creation and implementation of risk scoring and anomaly detection tools at the level of the program agency), and human capital (providing training for compliance staff to interpret and follow up on the results of analyses). Each of these three investments represents an upfront cost, but leads to long term cost savings through better regulatory detection and less avoidance failures.

Empirical Evidence: Moving Beyond Compliance Theater

A third principle of reform is that resource allocation should be monitored in accordance with empirical evidence, not administrative practice. Based on the FAC data, there is a distinct ranking of priorities with SNAP (96.09%), IHWAP (91.12%), CSBG (89.63%), and LIHEAP (85.45%) regularly exhibiting the highest finding rates and requiring the most targeted enforcement. Congress should require GSA to add structured fields to FAC submissions capturing corrective action status and management decision timeliness; OMB should require federal program agencies to report risk-based monitoring decisions and outcomes in their annual performance reports; and Treasury should expand USASpending.gov integration with FAC to enable real-time cross-reference of award data with audit finding status. These reforms require no new authority — only systematic application of existing data to oversight decisions agencies are already making.

COSO-Based Implementation: From Framework to Practice

The COSO Internal Control Framework (2013) and the GAO Green Book provide the normative infrastructure for implementing the risk-stratified compliance model we propose. Specifically, the COSO framework's five integrated components, Control Environment, Risk Assessment, Control Activities, Information and Communication, and Monitoring Activities, map directly onto the reform framework's requirements.

The Control Environment component, which establishes the tone and organizational culture within which internal controls operate, is foundational to compliance improvement in state agencies. The fact that jurisdictional and program-level patterns all match capacity explanations also provides some evidence that weaknesses in the control environment play a role in persistent noncompliance. It will not be sufficient to just attempt to meet the technical compliance requirements of the Uniform Guidance, but efforts will have to be made to invest in leadership for compliance, staff educational training, as well as ethical accountability procedures to address these weaknesses.

The Risk Assessment component requires agencies to identify and analyze risks to the achievement of compliance objectives, which is precisely the purpose of the risk scoring model proposed in Section V.B. This analytical basis for allocating monitoring resources in proportion to risk is systematic and done at regular intervals according to vulnerability thresholds relevant to each program.

The most glaring disconnect between the COSO framework and current compliance practice is in the Monitoring Activities component, the dynamic assessment of whether the components of internal control are functioning effectively. The Fiscal Control and Internal Auditing Act (FCIAA) in Illinois, as well as other states, requires annual internal control certifications, but these certifications take place through a static self-assessment process rather than allowing for any real-time monitoring capabilities. Incorporating the existing FCIAA certification process into continuous transaction monitoring would significantly strengthen the monitoring function without the need for new legislation.

Policy implications

Implications for Congress

Our findings support several specific legislative actions. First, Congress should modify Single Audit Act implementing requirements to mandate that federal program agencies develop and publish documented risk-based monitoring frameworks for major grant programs, with explicit criteria for risk stratification and monitoring intensity calibration. Currently, the regulatory requirement for risk-based subrecipient monitoring at 2 C.F.R. § 200.332 applies to pass-through entities but does not apply equivalent requirements to federal program agencies' monitoring of direct recipients. Closing this gap would extend risk-based monitoring principles to the full compliance oversight chain.

Second, Congress should require GSA to develop and publish a corrective action tracking system within the FAC that enables federal agencies, Congress, and the public to monitor the resolution status of Single Audit findings in real time. The GAO's (2024a) finding that 213 pre-2015 findings remained unresolved in 2021 reflects in part the absence of systematic resolution tracking, a gap that a relatively modest investment in FAC infrastructure could address.

Third, Congress should consider targeted capacity-building investments for state agencies administering federally funded programs with persistently elevated finding rates. Findings resulting in smaller jurisdictions and higher complexity programs are in line with capacity-related explanations, and indicate that technical assistance in compliance program design, data analysis and the implementation of internal controls may lead to more sustainable compliance improvements than enforcement alone can provide to organizationally constrained recipients.

Implications for OMB and Federal Program Agencies

For OMB, our findings support updating Circular A-123 and the Uniform Guidance implementation guidance to require federal program agencies to document the use of FAC finding data and improper payments estimates in subrecipient monitoring decisions. This requirement would ensure accountability for allocating resources based on proven results without specifying the exact methods of monitoring, allowing discretion at the program offices while setting a minimum standard of accountability for the use of evidence.

For programming agencies at the federal level, our findings would suggest investments in transaction-level analytics capabilities that could allow for continuous monitoring of high-risk grantees. The CMS Fraud Prevention System provides a model for federal continuous monitoring on a large scale; similar capabilities adapted to the social service and community development grant context would allow for a substantial increase in detection without increasing the compliance burden on the majority of recipients with consistently clean records.

Implications for State-Level Practitioners

Among state agencies that administer programs run with federal funds, the findings support three practical changes. First, internal audit functions at the state level should begin to develop risk scoring models for subrecipient portfolios using data on historical findings by the FAC, characteristics of the award, and indicators of organizational capacity. These models do not require machine learning infrastructure, a weighted scoring rubric using existing data is a defensible first step. That is, state agencies ought to incorporate transaction-level analytics into their regular compliance monitoring processes rather than only resorting to these methods on an ad hoc basis through the accrual of documentation every few months. Third, state agencies need to tie FCIAA and Green Book internal control certification processes to real-time monitoring capacity rather than have them serve as an annual paperwork exercise.

Policy Priority Table

Given the range of reforms proposed above, a clear hierarchy of priorities may assist policymakers and practitioners in sequencing action. Table 12 lists the reform recommendations by level of priority, actor, timeline for implementation, and level of expected impact on compliance outcomes.

Table 12: Policy Priority Framework for Federal Grant Compliance Reform

Priority	Reform	Actor	Timeframe	Expected Impact
Priority 1	Add corrective action status tracking to FAC	Congress / GSA	Short-term (1–2 years)	Enables real-time resolution tracking; directly addresses the GAO finding that 213 pre-2015 findings remained unresolved in 2021
Priority 2	Publish program-specific risk indicators on PaymentAccuracy.gov	OMB / Federal agencies	Short-term (1–2 years)	Supports evidence-based subrecipient monitoring resource allocation by all state practitioners
Priority 3	Require federal agencies to document risk-based monitoring frameworks	OMB / Congress	Medium-term (2–4 years)	Aligns federal oversight resources with empirically identified risk; reduces uniform audit burden on low-risk recipients
Priority 4	Fund state compliance analytics capacity-building	Congress / OMB	Medium-term (2–4 years)	Addresses organizational capacity as a driver of finding persistence; enables continuous monitoring at the state agency level
Priority 5	Align Single Audit follow-up with continuous monitoring	OMB / Federal program agencies	Long-term (4+ years)	Eliminates temporal monitoring gaps; reduces detection lag from 18–42 months to 3–6 months

Priorities 1 and 2 require minimal new resources and can be implemented within existing agency authorities. So, they are proposed as immediate first steps even as longer-term structural reforms continue to be worked toward.

These analyses lead to one concrete inference: violations of federal grants administered by the states are not random noise. They are concentrated and persistent, and they are predictable. The top 12.40% of uniquely identified entities make up 90% of all finding years. Previous failure to comply is predictive of future failure at both two and three years. Finding rates remained stable for eight years at just under 47%. These patterns are not arguments against reform; they are arguments for targeting it precisely. Risk stratification based on COSO principles, monitoring that is calibrated to the history of findings at the entity level, and mechanisms for taking corrective action that are aligned with data-driven assessments of risk rather than cycles of documentation would direct federally oversight resources exactly where the data show that they are needed most. The data that can make this happen already exist within the Federal Audit Clearinghouse, PaymentAccuracy.gov, and USASpending.gov. The framework to use them systematically is the contribution this paper provides.

CONCLUSION

This article studied compliance failure in federal grant programs run by the states. It used long-term data from the Federal Audit Clearinghouse, covering 7,147 state-program-year observations from 2016 to 2023. The results support the main argument of this study. Periodic accountability systems can document non-compliance, but they do not help organizations learn from it. H1 is strongly supported. Just 12.40% of entities account for 90% of all finding-years (Gini = 0.8902). H2 is also supported. The findings from two years earlier predict current persistence at 9.48 percentage points (p = 0.0002). H3 receives moderate support, since larger awards are linked to less severe findings. Finding rates stayed near 47% for eight years, with no sign of improvement. Five separate indicators support the idea of symbolic compliance, which gives strong support to H5. These findings add to three areas of theory: principal-agent theory, audit society theory and organizational learning theory. Together, they show that occasional monitoring does little to deter non-compliance, and that risk patterns persist over time. The article offers a reform proposal built on flexible monitoring, continuous data use and evidence-based resource allocation. The study is descriptive, not causal, and has some data limitations. Even so, its findings call for real reform to protect the people who depend on these programs.

Data availability statement

The data underlying this study are drawn entirely from publicly available government sources. Single Audit finding data were obtained from the Federal Audit Clearinghouse (FAC) via the FAC public API at api.fac.gov, which provides open access to all Single Audit submissions by nonfederal entities expending federal funds. Improper payment estimates were obtained from PaymentAccuracy.gov, maintained by the U.S. Office of Management and Budget. Federal award data were obtained from USASpending.gov. The analytical panel dataset constructed from these sources, along with the Python scripts used for data extraction, variable construction, and regression analysis, are deposited in a publicly accessible replication repository and available upon request from the corresponding author. No proprietary, confidential, or restricted data were used in this study.

REFERENCES

Ash, E., Galletta, S., & Giommoni, T. (2025). A machine learning approach to analyze and support anti-corruption policy. *American Economic Journal: Economic Policy, 17*(2), 162-198.

Association of Certified Fraud Examiners (ACFE). (2024). *Report to the nations: 2024 global study on occupational fraud and abuse*. ACFE.

Blume, L., & Voigt, S. (2011). Does organizational design of supreme audit institutions matter? A cross-country assessment. *European Journal of Political Economy, 27*(2), 215-229.

Brusca, I., Caperchione, E., Cohen, S., & Manes Rossi, F. (2016). *Public sector accounting and auditing in Europe: The challenge of harmonization*. Palgrave Macmillan.

Centers for Medicare and Medicaid Services. (2023). *Fraud Prevention System: Annual report to Congress*. CMS.

Committee of Sponsoring Organizations of the Treadway Commission (COSO). (2013). *Internal control, Integrated framework*. COSO.

Congressional Research Service. (2024). *Searching the Federal Audit Clearinghouse: A brief guide* (IF12965). Library of Congress.

Congressional Research Service. (2024). *Improper payments: Ongoing challenges and recent legislative proposals (*R48296). Library of Congress.

Cyert, R. M., & March, J. G. (1963). *A behavioral theory of the firm*. Prentice-Hall.

Donahue, J. D., & Nye, J. S. (Eds.). (2002). *Market-based governance: Supply side, demand side, upside, and downside*. Brookings Institution Press.

Herd, P., & Moynihan, D. (2018). *Administrative burden: Policymaking by other means*. Russell Sage Foundation.

Ferraz, C., & Finan, F. (2011). Electoral accountability and corruption: Evidence from the audits of local governments. *American Economic Review, 101*(4), 1274-1311.

Government Accountability Office (GAO). (2014). *Standards for internal control in the federal government (GAO-14-704G)*. GAO.

Government Accountability Office (GAO). (2017). *Single audits: Improvements needed in selected agencies' oversight of federal awards* (GAO-17-159). GAO.

Government Accountability Office (GAO). (2024a). S*ingle audits: Improving Federal Audit Clearinghouse information and usability could strengthen federal award oversigh*t (GAO-24-106173). GAO.

Government Accountability Office (GAO). (2024b). *Improper payments: Key concepts and information on programs with high rates or lacking estimates* (GAO-24-107482). GAO.

Government Accountability Office (GAO). (2025a). *Improper payments: Agency reporting of payment integrity information* (GAO-25-107552). GAO.

Government Accountability Office (GAO). (2025b). *Program integrity: Agencies and Congress can take actions to better manage improper payments and fraud risks* (GAO-25-108172). GAO.

Harrington, S. J. (2019). Predicting single audit findings: Evidence from the Federal Audit Clearinghouse. *Journal of Government Financial Management, 68*(1), 22-28.

HHS Office of Inspector General. (2024). *NIH did not consistently meet federal single audit requirements for extramural grants*. HHS OIG.

Knechel, W. R., & Salterio, S. E. (2016). A*uditing: Assurance and risk* (4th ed.). Routledge.

Krishnan, R., Yetman, M. H., & Yetman, R. J. (2018). Accounting data quality, governance, and cost of debt financing: Evidence from single audits of state and local governments. *Journal of Accounting Research, 44*(5),

867-897.
Levitt, B., & March, J. G. (1988). *Organizational learning. Annual Review of Sociology, 14*(1), 319-340.
Office of Management and Budget. (2021). *Requirements for payment integrity improvement (OMB Circular A-123, Appendix C, M-21-19).* OMB.
Power, M. (1997). *The audit society: Rituals of verification.* Oxford University Press.
Plummer, E., Raman, K. K., & Zheng, X. (2020). The effect of organizational characteristics on the likelihood of single audit findings. *Journal of Governmental and Nonprofit Accounting, 9*(1), 1-24.
Posner, P. L. (1998). The politics of unfunded mandates: Whither federalism? Georgetown University Press.
Romzek, B. S., & Dubnick, M. J. (1987). Accountability in the public sector: Lessons from the Challenger tragedy. *Public Administration Review, 47*(3), 227-238.
Romzek, B. S., & Ingraham, P. W. (2000). Cross pressures of accountability: Initiative, command, and failure in the Ron Brown plane crash. *Public Administration Review, 60*(3), 240-253.
Vasarhelyi, M. A., & Halper, F. B. (1991). The continuous audit of online systems. Auditing: *A Journal of Practice and Theory, 10*(1), 110-125.
Wang, W., & Vasarhelyi, M. A. (2024). The application of continuous audit and monitoring methodology: A government medication procurement case. *International Journal of Accounting Information Systems, 55*, 100713.
Waterman, R. W., & Meier, K. J. (1998). Principal-agent models: An expansion? *Journal of Public Administration Research and Theory, 8*(2), 173-202.
Wood, B. D., & Waterman, R. W. (1991). The dynamics of political control of the bureaucracy. *American Political Science Review, 85*(3), 801-828.
Wright, D. S. (1988). *Understanding intergovernmental relations* (3rd ed.). Brooks/Cole.

Robustness Checks for H2 (FindingPersistence): Addressing Definitional Overlap

FAC Panel Data 2016–2023 | State-Program-Year Observations

Overview and Motivation

The primary specification for H2 in Table 8 of the main text uses Prior Year Finding (t-1) as the key predictor of Finding Persistent (which equals 1 if the entity received a finding in both year t and year t-1). A reviewer correctly noted that this specification contains a partial definitional overlap: because Finding Persistent requires a finding in year t-1 by construction, any observation with Prior Year Finding = 0 is mechanically guaranteed to have Persistence = 0. This is confirmed by the separation diagnostics in Table A.1 below.

This appendix addresses the concern in two ways. Table A.1 documents the severity of the separation to show why standard logit is inappropriate and why the LPM is the correct primary specification. Table A.2 reports robustness specifications using lagged predictors from t-2 and t-3, which are definitionally independent of the outcome. Those specifications provide the cleanest test of whether genuine organizational persistence exists beyond the definitional structure of the t-1 model.

Separation Diagnostics

Key implication

The t-1 specification is not a valid test of whether compliance failure reflects stable organizational traits. When Prior Year Finding = 0, Persistence = 0 by construction, not because the entity has no persistent risk. The t-2 and t-3 specifications break this mechanical link. If persistence survives there, the underlying pattern is real.

Table A.1: Separation Analysis: Prior Year Finding and Persistence by Cell

Prior Year Finding	N Observations	Persistence = 0 (N)	Persistence = 1 (N)	Persistence Rate
0 (No prior finding)	2,936	2,936	0	0.00%
1 (Prior finding present)	2,781	659	2,122	76.38%
Total	5,717	3,595	2,122	37.12%

Note: Source: FAC, 2016–2023. N = 5,717. Sample excludes 2022 (UEI transition anomaly) and one Texas outlier. The diagnostic confirms perfect (not merely near-perfect) separation in the t-1 specification: zero observations have Persistence = 1 when Prior Year Finding = 0. This follows from the variable construction and confirms that standard logit is inappropriate for this specification. The LPM in Table 8 remains valid as a primary specification because it does not assume a specific functional form for the probability. The 76.38% conditional persistence rate (the share of entities with a prior finding that received another finding the following year) is a descriptive statistic unaffected by this issue.

Robustness Checks Using Lagged Predictors (t-2 and t-3)

To address the definitional overlap, the following specifications replace the immediate lag (Prior Year Finding, t-1) with a two-year lag (Prior Finding t-2), a three-year lag (Prior Finding t-3), and a three-year rolling sum of findings from t-2 through t-4. These predictors cannot share the definitional relationship with the current outcome. All specifications include program, state, and year fixed effects with standard errors clustered at the state level.

Table A.2: LPM Robustness Specifications — Lagged Predictors Addressing Definitional Overlap

Specification	N	Lag Predictor	Coefficient	Std. Error	p-value	R^2
Model 1: Original t-1 lag (Main Table 8)	5,717	t-1	0.4908***	(0.028)	< 0.001	0.7695
Model 2: Two-year lag (Definitionally clean)	5,203	t-2	0.0948***	(0.026)	0.0002	0.6206
Model 3: Three-year lag	5,162	t-3	0.0757***	(0.019)	0.0001	0.6158
Model 4: 3yr rolling sum from t-2 to t-4	5,203	t-2 to t-4	0.0629***	(0.012)	< 0.001	0.6243
Fixed Effects (all models)	—	—	Program, State, Year			
Clustered SE	—	—	State level (all models)			

*Note: Source: FAC, 2016–2023. Dependent variable: Finding Persistent (binary). All models are OLS linear probability models. *** p < 0.001, ** p < 0.01, * p < 0.05. Standard errors clustered at state level in parentheses. The t-1 lag (Model 1) is reproduced from Table 8 for reference. Models 2–4 use predictors that are definitionally independent of the current outcome because they do not share the t-1 construction. The reduction from N = 5,717 to N = 5,203 in Models 2 and 4 reflects the loss of the earliest observation year per entity-program combination when constructing the two-year lag.*

Interpretation

The t-2 and t-3 results are clear on the key question. The finding two years prior is associated with a 9.48 percentage point increase in current persistence (SE = 0.026, p = 0.0002). At t-3, the effect is 7.57 percentage points (SE = 0.019, p = 0.0001). Across the rolling three-year sum from t-2 through t-4, the coefficient is 6.29 percentage points (p < 0.001). All three clean specifications are statistically significant, confirming that genuine organizational persistence exists independent of the definitional issue.

The drop from 49.08pp (t-1) to 6.29–9.48pp (t-2 and later) is expected and tells you something useful: the t-1 coefficient is mostly capturing the definitional constraint, not organizational risk. The reviewer was right about that. But the residual effect at t-2 and t-3, while smaller, is real and consistent across multiple lag specifications. Entities with prior compliance failures are systematically more likely to fail again, even when that relationship is measured two and three years removed from the outcome.

The honest characterization of H2 is this: the 76.47% conditional persistence rate is partly a product of variable construction, and the 49.08pp regression coefficient combines a definitional component with a genuine organizational persistence component. The genuine component, isolated at t-2, is 9.48 percentage points. That is smaller than the main table implies, but it is statistically significant and survives three different lag specifications. H2 holds, more modestly than originally claimed.

Note on logit specifications: Table A.1 confirms perfect separation in the t-1 binary logit specification (zero observations with Persistence = 1 when Prior Year Finding = 0). Perfect separation means standard logit does not converge reliably regardless of penalty choice. Firth penalized logit requires true Jeffreys-prior penalization, which is implemented in R's logistf package but is not yet available in Python's statsmodels. The Python regularized logit (L1/L2 penalty) is not a true Firth implementation and did not converge in testing. For this reason, the LPM is the correct and preferred primary specification for this outcome, and logit robustness checks are appropriately deferred. Researchers wishing to replicate a Firth logit check may use the replication dataset with R's logistf package: logistf(finding_persistent ~ prior_finding + log_expended + n_programs + factor(program) + factor(state) + factor(year), data = panel_clean, firth = TRUE, pl = TRUE)

Appendix B: Supplementary Tables

Table B.2: Regression-Based Evidence Summary, All Hypotheses

Hypothesis	Revised Statement	Empirical Evidence	Status
H1: Concentration	Small subset of entities accounts for most findings	Top 12.40% = 90% of finding-years; Gini = 0.8902 (Table 5)	Strong support
H2: Persistence	Prior finding history predicts future persistence	LPM t-1: +48.75 pp (p < 0.001); clean t-2 estimate: +9.48 pp (p = 0.0002) (Table 8, Appendix A)	Strong support
H3: Capacity	Capacity differences associated with finding rates and severity	Award size: coef = -0.438 on severity (p < 0.001); territorial patterns consistent with capacity (Tables 6 and 9)	Evidence broadly consistent with proposition

System-Level Monitoring Proposition	Finding rates show no sustained decline	Stable at 43–50% across 2016–2023; monitoring not directly measured (Table 3)	Descriptive system-level evidence
H5: Compliance Theater	Post-finding trajectories do not show sustained improvement below baseline	76.47% persistence given prior finding; post-finding rates return to 47% baseline not below (Tables 4 and 10)	Strong descriptive support; formal causal test pending

Note: This table integrates descriptive and regression-based evidence. It is reproduced in full in Appendix B for reference.

Table B.1: Finding Rates by State and Territory, Top 20

State/Territory	Obs.	Finding Rate	Severe Rate	Persistence Rate
AS† (American Samoa)	10	1.0000	1.0000	0.8000
DC† (District of Columbia)	56	1.0000	1.0000	0.4821
GU† (Guam)	27	1.0000	1.0000	0.7037
MP† (N. Mariana Islands)	24	1.0000	1.0000	0.7083
VI† (Virgin Islands)	31	1.0000	1.0000	0.5806
NV	64	0.9531	0.9531	0.8125
MD	84	0.9167	0.9167	0.6786
NH	60	0.9000	0.9000	0.8000
NE	67	0.8657	0.8657	0.6418
RI	71	0.8592	0.8592	0.6056
VT	58	0.8448	0.8448	0.5862
HI	74	0.8243	0.7973	0.6622
SD	74	0.8108	0.8108	0.5135
DE	58	0.7931	0.7931	0.5172
MS	76	0.7895	0.7895	0.5658
MA	80	0.7875	0.7625	0.6500
MT	40	0.7750	0.7750	0.4000
CT	84	0.7619	0.7619	0.4881
VA	97	0.7423	0.7010	0.5876
WA	80	0.7375	0.7375	0.5625