BLAVATNIK SCHOOL OF GOVERNMENT

WORKING PAPER

PROVIDING ACCESS TO THE LATEST POLICY-RELEVANT RESEARCH

LOW-STAKES ACCOUNTABILITY AND PUBLIC SERVICE TURNAROUNDS

Thomas Elston, University of Oxford

Han Wang, University of Southampton

BSG-WP-2025-63

September 2025

Low-stakes accountability and public service turnarounds

Thomas Elston[1] (University of Oxford) and Han Wang (University of Southampton)

ABSTRACT

Can accountability to external bodies induce performance turnarounds in struggling public services? And if so, must account-holders use incentives and sanctions to change organizations that have yet to self-correct, or can a gentler, more informational regime suffice? Using recent research into "low-stakes" accountability, we argue that feedback and standard-setting on their own may stimulate subunit turnarounds in complex, multi-service organizations by directing leaders' scarce attention, control efforts and resource allocations toward previously unnoticed performance deficits. However, given the potential for "tunnel vision" among account-givers, accountability-induced turnarounds may be confined to dimensions of performance most relevant to the account-holder, to the neglect of others. We test each stage of this theory using quasi-experimental methods and data from the Local Government Ombudsman in England. We show that councils that are notified of maladministration in their social care provision significantly increase leadership

[1] thomas.elston@bsg.ox.ac.uk

attention and control efforts toward this service, and invest more in its core staffing, compared with a matched sample of unnotified councils. On average, the rate of maladministration falls by more than half; though, as predicted, we detect no wider performance gains. The immediacy of the improvement further suggests that responsible subunits may act *in anticipation* of leaders' increased attention.

KEYWORDS

Accountability; turnaround; difference-in-differences; ombudsmen; tunnel vision

ACKNOWLEDGEMENTS

This research was funded by an Engagement Fellowship from the Higher Education Innovation Fund (HEIF). We are grateful to staff from the Local Government and Social Care Ombudsman for providing data and advice, and for hosting Dr Wang's secondment during 2023. We also thank Rachel Darby, without whose excellent support our dataset would have been far smaller. And we are grateful for comments from Benny Geys.

1. INTRODUCTION

Ensuring and enabling the turnaround of underperforming public services is an important responsibility for policymakers, regulators and public managers alike (Jas & Skelcher, 2005; Meyers & Murphy, 2007; Boyne & Meier, 2009; Rutherford & Favero, 2020; Alonso & Andrews, 2021). At present, most research into organizational turnarounds focuses on troubled firms (see reviews in Trahms et al., 2013; Schweizer & Nienhaus, 2017; Castelló-Sirvent & Roger-Monzó, 2023), although there is a good track record of adapting these insights for the public sector (Boyne, 2006; Harvey et al., 2010; Murphy, 2010). Yet public service turnarounds also differ in that they typically take place under the "glare of public accountability."

The likely effect of accountability to external bodies on the speed, quality and sustainability of public service turnarounds is at present unclear. Psychologists have shown in lab experiments that judgement tends to improve under pressure of accountability (Schillemans, 2016; Aleksovska et al., 2019). Some econometric studies of school and hospital accountability also report positive effects (Kelman & Friedman, 2009; Propper et al., 2010; Dee & Jacob, 2011; Burgess et al., 2013), although others are more cautious (Deming et al., 2016; Hofer et al., 2020). And the public administration literature on accountability lists many cases of oversight regimes that failed to deliver performance gains, consumed significant organizational resources, and/or encouraged various dysfunctional behaviors (Romzek & Dubnick, 1987; Dubnick, 2005; Bevan & Hood, 2006). Thus, the evidence base relating

accountability to public service improvement *in general* is mixed (Dubnick, 2005), even before its applicability to struggling agencies awaiting turnaround is considered (see Rutherford, 2014).

Different *types* of accountability may also affect turnaround prospects differently. Intuitively, it might be reasoned that "the worse the malady, the stronger the medicine," in which case *high*-stakes accountability will be needed to change organizations that have hitherto failed to self-correct. This means imposing costly sanctions for continued failure/or and tangible rewards for improvement, as judged by the account-holder. And yet, recent research into gentler performance regimes suggests that feedback and standard-setting alone, without significant economic or reputational consequences, might be more suited to fostering the deep and reflective organizational learning necessary to improve public services (Jakobsen et al., 2018; Andersen & Nielsen, 2020). In particular, dysfunctional behaviors like threat rigidity and gaming may lessen under such *low*-stakes regimes, while open debate and broad search activity may increase. However, empirical tests of low-stakes accountability remain rare (examples are Woo et al., 2015; and Andersen & Nielsen, 2020); and, again, it is unclear whether this approach is suited to controlling agencies that are particularly struggling.

To advance this debate, we synthesize the literatures on turnaround and low-stakes accountability to theorize and then test the effects of feedback and standard-setting alone on underperforming public services. We focus on complex, multi-service organizations (like local governments) in which a

variety of activities are undertaken by distinct organizational subunits under common leadership. We argue that initiating turnarounds in subunits will depend on external accountability bodies directing leaders' limited "budget" of attention toward performance deficits that were hitherto unnoticed or downplayed. Absent the threat of financial or reputational consequences for inaction, leaders will need to receive novel yet credible evidence of service failure sufficient to disconfirm their prior beliefs (see Meier et al., 2015; Andersen & Nielsen, 2020). When this occurs, they may increase control over the problem areas and divert resources from peripheral to core activities – both of which are recognized turnaround strategies (Boyne & Meier, 2009; Murphy, 2010; O'Kane & Cunningham, 2014). However, given the risk of "tunnel vision" among any account-giver (even in low-stakes regimes), performance recovery in the subunit may still be concentrated on dimensions within the remit of the external account-holder, to the neglect of other "extra-jurisdictional" aspects. That is, accountability-induced turnarounds will be inherently *partial*.

We test these hypotheses quantitatively using quasi-experimental methods and the case of the Local Government Ombudsman in England. This is a prime example of a seemingly underpowered accountability body tasked with informing and advising but not sanctioning or coercing its account-givers (Kirkham, 2005; Thomas, 2024; see also Hertogh, 2001 and Harlow, 2018). (Indeed, "ombudsman or ombud*smouse?*" is a question periodically asked of

these institutions, so feeble are their powers of enforcement believed to be.)[2] Using coarsened exact matching and staggered difference-in-differences, we show that councils that are notified by the Ombudsman of maladministration in their adult social care provision significantly increase leadership attention and control efforts toward this service, and invest more in its core staffing, compared with a matched sample of untreated councils. The incidence of maladministration also falls in notified councils by, on average, 57% (equivalent to 3.05 fewer Ombudsman-upheld complaints per year). However, consistent with the expected tunnel vision, we detect no broader performance gains on a wide variety of additional measures. Moreover, the immediacy of the improvement suggests that, once notified of failures, subunits proactively seek to recover performance *in anticipation* of leaders' increased attention, rather than in its aftermath. This is a novel finding in turnaround literature.

The article proceeds as follows. The second section predicts the effects of low-stakes accountability on performance turnaround, and the third describes our empirical case. The fourth section presents the data and empirical strategy, the fifth provides the results and robustness checks, and the sixth discusses the findings.

---

[2] Most notably, this was the title of an article by William Gwyn (1973) about the ombudsman for national government – whose statutory makeup is very similar to the LGSCO, but who at least benefits from the support of a dedicated parliamentary committee (see Harlow & Rawlings, 2021, ch. 13).

2. ACCOUNTABILITY AND TURNAROUNDS

2.1 Low-stakes accountability, learning, and tunnel vision

For many commentators, it is imperative that accountability regimes are capable of imposing meaningful consequences on account-givers if they are to induce behavioral change and performance improvement. In the regulation literature, for instance, the principle of the "benign big gun" rests upon regulators having access to credible and costly sanctions against wrongdoers, even if these are only to be used sparingly (Ayres & Braithwaite, 1992). Moreover, one of the main challenges in effectively regulating public-sector entities (compared with firms) is thought to be the reticence of one part of the state to litigate against another (Wilson & Rachal, 1977; Konisky & Teodoro, 2016). In administrative law, too, the dominant view is that "the effectiveness of the courts and the ombudsman is primarily determined by the legal force [or not] of their decisions" (Hertogh, 2001). And similarly, in accountability studies, it is often argued that "accountability cannot flourish if the forums, the account holders, are weak" (Bovens & Wille, 2021, p. 857). As for performance management research, while there is growing interest in using non-financial means of reward and sanction, the presumption remains that some form of meaningful consequences must be tied to results. As Bevan (2010, p. 36) argues: "whatever system of performance measurement is used, to have an impact it needs to be designed to inflict reputational damage on poor performers."

On the other hand, there is also some emerging interest in the potential for low-stakes accountability to drive performance improvement in public services (Woo et al., 2015; Jakobsen et al., 2018; Andersen & Nielsen, 2020). "Low-stakes" denotes regimes where the measurement and judgement of performance proceed on the same basis as in high-stakes systems, but without any consequences being tied to those findings (see Heinrich & Marschke, 2010; Verger & Parcerisa, 2017). That is, there are no automatic or probable economic effects (for individuals or organizations), reputational incentives (for instance, use of rankings or league tables, which may have second-order economic effects), or other forms of top-down intervention (like merging or closing inadequate agencies) tied to the judgement of the account-holder.

In part, interest in low-stakes accountability arises from concern in public administration and allied fields (like education and healthcare) over the track record of using high-stakes regimes for public service improvement, particularly in terms of undesirable side-effects. As Jakobsen et al. (2018, p. 127) write, "it is far from certain that the net impact … has been positive" (although see Gerrish (2016) and Boyne (2010) for more optimistic assessments). The hope is that if gaming or other pathologies stem mainly from the rewards-and-sanctions elements of accountability, then removing high-powered incentives should tip the balance in favor of performance gains (de Wolf & Janssens, 2007; Dee & Jacob, 2011; Verger & Parcerisa, 2017; Jakobsen et al., 2018; although see Levatino et al., 2024 for evidence

refuting the idea of significant differences in opportunism between high- and low-stakes regimes).

Beyond attenuating negative-side effects, low-stakes regimes may also improve the account-giver's capacity for learning. As Jakobsen et al. (2018, p. 135) write, "goal-oriented learning and innovation will become more likely when organizational actors perceive that they work in a setting where acknowledging errors and problems will lead to a dialogue about problem solving, rather than punishment." Similarly, Andersen and Nielsen (2020, p. 416) argue that, with incentives removed, novel performance information allows actors to "update their prior beliefs … and use these new beliefs to prioritize their efforts." Crucially, Andersen and Nielsen do not argue that *motivation to learn* does not matter, but rather that intrinsic public service motivation may provide sufficient inducement, once the inadequacy of current performance is understood. This contrasts markedly with the "worse malady, stronger medicine" argument, and signals that low-stakes accountability adopts more of a "trust" than a "mistrust" approach to governing public services (Le Grand, 2010).

The mechanisms expected to enhance organizational learning are not well specified in this literature, although there are several possibilities. One is that high-stakes typically also means high-*stress* for staff and managers, which may impede learning processes. For instance, research on "threat rigidity" in organizations suggests that channels of information and means of adaptation become constricted in the face of adversity, triggering old,

habitual and potentially ill-fitted behaviors (Staw et al., 1981). Another possibility is that high-stakes regimes divert scarce resources toward unproductive activities, like temporary fixes or superficial "window dressing" (Smith, 1995), resulting in less capacity for genuine search and innovation. And a third is that high-stakes accountability may restrict the scope of any learning to domains linked to rewards or penalties, contributing to what Moynihan (2005, p. 203) calls "narrow process improvement … rather than broad understanding of policy changes."

Finding ways to enhance organizational learning is especially relevant to questions of turnaround, because failing to detect and respond to informative signals is one of the chief contributors to decline. As McKiernan (2006, p. 771) explains:

> "Primary causes of decline are the result of defects in the corporate learning process. … Companies with good learning routines will normally be expected to survive … Companies with defective routines are more likely to struggle and so to drift down the spiral of decline."

Many scholars of public sector turnaround adopt the same view (Boyne, 2006; Harvey et al., 2010; Harvey et al., 2015; Vindrola-Padros et al., 2022). Consequently, if an accountability regime is to help arrest public service decline and initiate turnaround, it needs to foster rather than further deplete the account-giver's capacity for learning. This means that accountability *without* consequences may be particularly suited for supporting turnaround. Thus, we hypothesize:

H1: Performance improves in organizations that are *notified* of failures, compared with similarly-performing organizations that are unnotified.

Nevertheless, lowering the stakes in accountability regimes is unlikely to suppress all undesirable effects of account-giving (Levatino et al., 2024). Low-stakes is not no-stakes. And even just singling out an activity for measurement can lead to effort prioritization, as implied by both Campbell's Law[3] and Goodhart's Law,[4] and by the familiar dictum that "what gets measured gets managed." This leads to the pathology known as "tunnel vision," where managers "focus on the dimensions of performance that are included in the measurement system, displacing other important but unmeasured aspects" (Mannion & Braithwaite, 2012, p. 570; see also Smith, 1995). In the context of turnarounds, tunnel vision would produce narrowly-focused recoveries, extending only so far as the remit or priorities of the account-holder extend. This may be especially likely for highly-specialized accountability bodies that provide narrow rather than broad-based feedback. Thus, we hypothesize a second main effect:

[3] "The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor." (Campbell, 1979)

[4] "Any observed statistical regularity will tend to collapse once pressure is placed upon it for control purposes" (Goodhart, 1981).

H1a: Among organizations notified of failures, performance improves more on dimensions that are within the accountability body's remit, and less on dimensions that are extra-jurisdictional.

Having derived our primary hypotheses, we now consider some of the behavioral responses we might expect following notification of failure.

2.2 Leadership attention

Much research into turnarounds suggests that recognizing the need for corrective action is the first hurdle needing to be overcome. Leaders typically have demanding jobs and limited attention spans (Simon, 1997; Tuggle et al., 2010), and frequently suffer from information overload (Walgrave & Dejaeghere, 2017). What is more, transmission of information to the top of hierarchies may be imperfect and delayed (Argyris, 1977), so that even attentive leaders can be poorly informed about performance among organizational subunits. All of this contributes to what Murphy and Jones (2016) call a "lack of self-awareness" among decision-makers, with the implication that that some kind of shock or "trigger point" is needed to initiate problem recognition and recovery efforts (Gopinath, 2005).

The situation is further complicated in the public sector by the contested nature of success and failure (Brewer, 2006; Andersen et al., 2016). Turnaround involves "a recovery in performance after a period of organizational failure" (Boyne & Meier, 2009, p. 835). For firms, this manifests as decline in market-share or financial return (see Pearce & Robbins, 1993). But

in government, performance is multidimensional and the moment of outright failure may be unclear (Paton & Mordaunt, 2004; Van de Walle, 2016). Persistent under-performance does not risk bankruptcy, and government services must be provided in even the most unfavorable circumstances (Hargrove & Glidewell, 1990). Consequently, it is not just leader attention that needs to be secured, but also problem acceptance. Leaders may need to be persuaded that the errors reported are serious and distinctive among comparable organizations – rather than being "par for the course," a product of difficult operating circumstances, or a function of government's "impossible jobs" (see Hargrove & Glidewell, 1990; Entwistle & Doering, 2024).

One way to accomplish this reset is to provide feedback that is both novel yet highly credible (Andersen & Nielsen, 2020). If the account-holder can convince leaders that prior beliefs about performance in part of their organization were misplaced or outdated, the resulting "gap" between expectations and new evidence may be sufficient to initiate remedial actions (Meier et al., 2015). Of course, the more the feedback disconfirms prior beliefs, the more skeptical leaders may be as to its veracity, and the harder the account-holder must work. Leaders may be tempted to dismiss problems as "one-offs" or "noisy data" – leading to what McKiernan (2006) calls "ostrich" behavior. Hence, turnaround is more likely when account-holders have strong claims to the credibility of both their data and their interpretation of it as falling below minimum performance standards. When this is the case, we expect:

H2: Following notification of failures, leaders pay more attention to the problematic service area than they did prior to notification.

2.3 Centralization, core competencies and urgency

With leader attention and problem acceptance secured, various internal turnaround strategies may be instigated (see reviews in Boyne, 2006; Trahms et al., 2013; Schweizer & Nienhaus, 2017; Castelló-Sirvent & Roger-Monzó, 2023). For now, we focus on centralization of control over the problem, diversion of organizational resources from peripheral to core activities, and prompt implementation.

Firstly, when leaders are concerned about performance problems, they may seek greater control over the subunit responsible for the errors (O'Kane & Cunningham, 2014; Meier et al., 2015), at least until improvement takes hold. As O'Kane and Cunningham (2014, p. 965) write, "'Hard' leadership is important during turnaround to centralize command, reduce participation and instill close supervision and control." McKiernan (2006) summarizes this approach as "more autocracy." The point is not only that sub-unit discretion declines, but also that enhanced efforts are made to ensure that commands issued at the top are obeyed below – lest recovery be impeded by non-adherence to the turnaround strategy. Thus, we hypothesize:

H3: Following notification of failure, leaders will temporarily centralize control over the responsible subunit.

Secondly, successful turnarounds often involve a shift in resources from peripheral to core purposes. The idea is to dispense with inessential (or perhaps not-*imminently* essential) activities, and redirect staff and managers toward "doing what they know best" (McKiernan, 2006, p. 760) and "attending to the basics … to the work that really counts" (Murphy, 2010, p. 169). Practically, as Boyne and Meier (2009, p. 859) write, this often means "replacing non-core with core staff." Thus, we hypothesize:

> **H4:** Following notification of failures, organizations invest more in core competencies in the responsible subunit.

Finally, just as "time is of the essence" in successful corporate turnarounds (Tangpong et al., 2015; Barbero et al., 2017), we similarly expect that more rapid reactions by public managers will produce larger recoveries. In firms, the aim is to stem decline before the loss of market share and/or investor confidence creates a downward spiral in the organization's resource base. In government, analogously, the sooner the agency enters a recovery phase, the easier it will be to retain critical staff unnerved at the downturn, recruit new high-quality staff who would otherwise be deterred from "joining a sinking ship," retain good will from both service users and overseers, and escape top-down regulatory intervention. Hence:

> **H5:** Following notification of failures, organizations that implement changes more rapidly will experience a greater recovery in performance.

2.4 Summary of theory

To summarize, our primary proposition is that external accountability *can* induce performance turnaround in failing organizations (or subunits thereof), but that a low-stakes regime may be sufficient if it enhances organizational learning and lessens unproductive gaming. However, absence of high-powered incentives is unlikely to fully suppress tunnel vision, so we also expect accountability-induced turnarounds to be partial rather than comprehensive, focused on dimensions of performance most relevant to the account-holder. To initiate public service turnarounds, leadership attention needs to be captured through novel but credible performance feedback and authoritative standard setting. When this occurs, internal turnaround strategies will begin, including centralized control over problems and invest more in core competencies. The more rapidly this occurs, the greater the recovery.

3. EMPIRICAL CASE

3.1 The Local Government Ombudsman

The Local Government and Social Care Ombudsman[5] is responsible for investigating complaints from users of local public services in England, and for encouraging lesson learning and service improvement. Led by a named Ombudsman, the LGSCO employs nearly 200 staff and is one of several third-party complaint handlers in the UK's administrative justice system (see Gill, 2018; Harlow & Rawlings, 2021; McBurnie et al., 2023). The LGSCO can only commence investigations after efforts at resolving the complaint locally have failed, and its remit extends only to matters of maladministration and service failure.[6] Broader issues of policy and performance cannot be investigated. Maladministration concerns "the manner in which decisions are reached," and whether any bias, neglect, inattention, delay, incompetence, ineptitude, perversity, turpitude, arbitrariness, or similar was involved (Parpworth, 2022, p. 391). Service failure typically means failure to provide a

[5] In statute, the Commission for Local Administration.

[6] This remit resembles the issues studied by the "service recovery" literature (Thomassen et al., 2017; Döring, 2022; Caillier, 2023), although complaints referred to Ombudsman tend to be more serious and are always subject to some form of dispute between citizen and provider.

service to which the complainant was entitled. For simplicity, hereafter we refer to both dimensions as just "maladministration."[7]

On average during 2012-2019, the LGSCO received 11,772 complaints per year, undertook 5,947 investigations, and issued 3,483 adverse findings per year in which the complainant was found to have suffered injustice following maladministration. With each upheld complaint, the LGSCO makes recommendations for how the council should respond, such as apologizing, providing services, improving procedures or staff training , and/or offering financial compensation (LGSCO, 2024b). Anonymized decisions and recommendations are published online, but the LGSCO has no powers to compel councils to act upon them (Kirkham, 2005; Thomas, 2024). In addition, in a very small minority of cases, the Ombudsman herself (rather than the organization at large) issues a "public interest report" to formally notify council leaders of the seriousness of the errors uncovered. This is mostly[8] a discretionary action taken in cases of recurring faults, significant injustices, systemic problems, frequent complaints about a single issue, or significant topical issues (LGSCO, 2024). Again, there are no powers to compel changes based on the report, although it must be discussed at a senior decision forum

[7] Indeed, when service failure was added to the legislation governing the Ombudsman in 2007, this largely formalized the LGSCO's existing approach under the flexible concept of "maladministration" (see Seneviratne, 2008).

[8] A report is also the legal default outcome where a council refuses to carry out the recommendations and the Ombudsman decides to maintain her position.

and notified to the local press (Thomas, 2024). Over time, the proportion of cases escalated in this manner has reduced, reflecting resource constraints and a desire to conserve the distinctiveness of this solitary tool of enforcement (Kirkham, 2005). Indeed, as our matching analysis below reveals, a substantial number of councils display observable characteristics associated with the escalation criteria, even though relatively few receive censure in any given year.

Ombudsmen provide an ideal case for exploring the effects of low-stakes accountability. As noted, the advisory nature of their work and lack of enforcement powers are widely recognized (Stacey, 1971; Gwyn, 1973; Harlow, 2018). And yet there is also a small body of qualitative work pointing to the remarkable influence of these "soft power" institutions on both complainant outcomes and broader organizational behavior (Hertogh, 2001; Kirkham et al., 2008; McBurnie et al., 2023; Thomas, 2024). As for the LGSCO specifically, this is clearly a low-stakes accountability regime. As Thomas (2024) writes, "of all UK ombuds, the LGSCO perhaps has the weakest structural position as regards compliance." It cannot compel changes in councils, nor sanction funding or autonomy, and nor does it call for disciplinary action against individual officials. We have seen no evidence of LGSCO findings being cited by central government or others to justify regulatory action. And our own analysis (Figure 8 below, and Table A8 in the appendix) shows no effects of public interest reports on leadership turnover in councils – whereas, historically, England's hospital rankings led to several

chief executives losing their jobs (Bevan, 2021), and headteacher retention rates are lowest in schools receiving the worst, "inadequate" judgement from the English education inspectorate (Lynch et al., 2017)).[9]

Most legal scholars regard the combination of publicity and moral authority as the principal means of potential impact for the LGSCO. Yet, while the Ombudsman's work is published, the prospect of significant reputational damage is remote. Benchmarked against Hibbard et al.'s (2003) four criteria for gauging the strength of reputation-based accountability systems, this regime appears very weak. LGSCO does *not* (i) rank organizations best-to-worst, as happens with school league tables; and its findings are *not* (ii) widely disseminated, given the decline in local media outlets and readership (Clark, 2021). Moreover, neither the ombudsman system in general, not the verdict of "injustice cause by maladministration," is (iii) clearly understood by the general public[10] – quite the reverse, in fact (Kirkham, 2005; Dunleavy et al., 2010). And nor is (iv) performance re-assessed to a regular schedule. Consequently, while external reputation may partially influence council deliberations (see McBurnie et al., 2023), there are far stronger means of

[9] Inspection can also be highly traumatic for school leaders, and in recent years contributed to at least one suicide (Perryman et al., forthcoming).

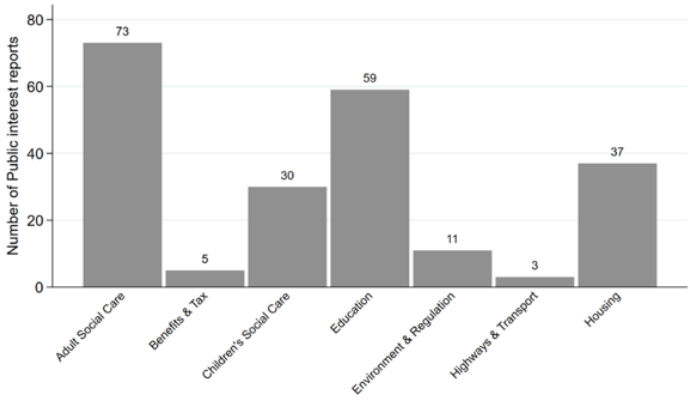
[10] Moreover, Hodges (2018, p. 54) suggests that lawyers and judges are often unfamiliar Ombudsmen systems, too.

leveraging reputation for purposes of accountability and performance improvement that are unavailable to the LGSCO.

3.2 Adult social care

We focus on the Ombudsman's oversight of councils' adult social care departments, partly because of the serious consequences of maladministration in this field, and partly because of the empirical opportunities that this provides. As Gulland (2024, p. 203) argues, "much social care decision making happens when people are at their most vulnerable and where outcomes can make life-changing differences to people's health and wellbeing." As such, while we do not claim to match the severity of performance failure examined in some turnaround research, maladministration in this domain "range[s] from the petty to the tragic" (Gulland, 2010, p. 457) and at times involves very serious individual and social harms. In addition, adult social care provides an excellent empirical opportunity for quantitative research. As Figure 1 shows, between 2012 and 2019 (our study period), LGSCO issued a total of 237 public interest reports to councils, of which adult social care accounted for the most (73 reports), followed by education (59 reports) and housing (37 reports). This is based on 6,955 upheld complaints about councils' adult social care over the period, amounting 28% of all upheld complaints. This provides ample scope for quantitative analysis.

Figure 1: Public interest notifications by service category, 2012-2019



Adult social care services involve the provision of personal care and domestic assistance to individuals aged 18 and over, including the elderly and others experiencing physical or mental illness, disability, or other challenging circumstances (Glasby et al., 2021; Burn et al., 2024). Unlike healthcare in England, social care is a local government responsibility and is means-tested (Needham & Hall, 2023). It is one of the most high-profile areas of work for English councils, and accounts for about a fifth of local public expenditure. This is rising with increasing life expectancy, an aging population, and the growing proportion of adults with diagnosed disabilities.

In England, care services are almost entirely *produced* by private or third-sector agencies (Goodair et al., 2024). (Complaints made against producers can also be investigated by the LGSCO, but are excluded from our analysis.) The primary role for councils is, therefore, to undertake the needs-testing and means-testing required by law to determine what level of support is required and how it should be funded (Needham & Hall, 2023; Burn et al., 2024). Most complaints to the Ombudsman relate to "delays in the assessment of

people's needs, processes that fail to put the individual at the center, care that fits with the system's offer rather than the person's needs and preferences, and poor communication" (LGSCO, 2024a, p. 1). In addition, councils have a broad well-being duty to promote quality of life among service users, a responsibility to ensure the availability of quality services in their areas, and a duty to support informal care givers like family members (Burn et al., 2024).

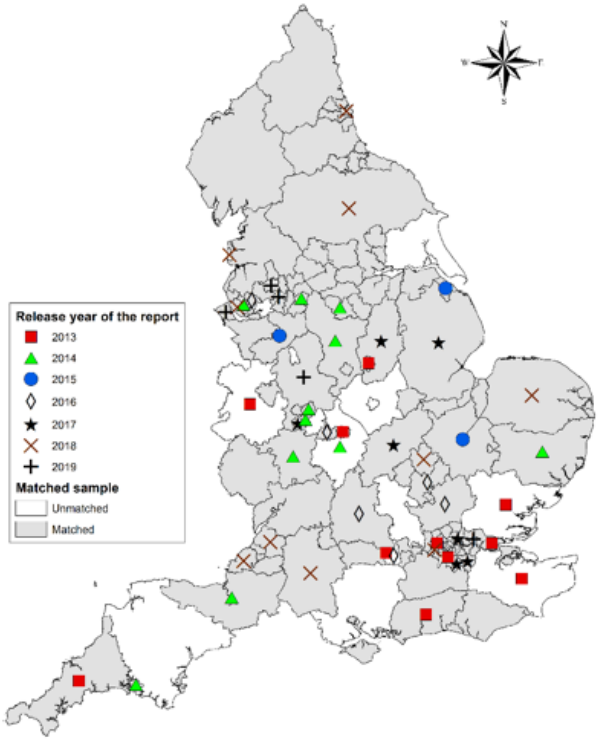
4. DATA AND METHODS

4.1 Variables and data sources

To test the effects of the Ombudsman's public interest notifications on turnaround outcomes and behaviors in adult social care, we construct a panel dataset at the upper-tier council level from 2012 to 2019.[11] We begin with public interest reports released in 2013, so that all analyses have at least one year of pre-treatment data; and we halt our analysis in 2019 to avoid biases due to COVID-19, which hugely disrupted this sector. Our variable of interest – the release of a public interest notification – is a dummy variable derived from the Ombudsman's online archive. For any given council, the variable is set to 1 for the year the report was issued and all subsequent years, and 0 otherwise. During our study period, 53 out of 151 upper-tier councils received at least one such a notification, with 12 receiving two reports, and four councils receiving three reports. Figure 2 illustrates the timing and geographic distribution of this low-stakes accountability "treatment."

[11] At the time of our analysis, local government in England consisted of five council types in two vertical arrangements: A two-tier system in predominantly rural areas, comprising 27 county councils (upper tier) and 201 district councils (lower tier); and single-tier authorities elsewhere, including 32 London boroughs, 36 metropolitan districts, and 55 unitary authorities. Social care is a responsibility of all council types except districts.

Figure 2: Release year of public interest report for each local authority



Regarding our dependent variables:

We quantify performance turnaround (H1) by examining the **incidence of maladministration** in adult social care departments overtime. This is determined by the rate of upheld complaints within the ombudsman's regular casework.[12] To test the breadth of any turnaround (H1a), we further analyze administrative data describing the speed of assessment activity, and two user satisfaction surveys. Specifically, we measure **the number of days of delay** in transferring individuals from hospital to adult social care due to

[12] This was supplied to us directly from the LGSCO as a case-level database, although all individual decisions can be inspected online for several years after determination. All survey data reported in this paragraph are also publicly available via government or NHS websites.

delays in care needs assessment or public funding (separately), or because of disputes between agencies. In addition, we employ **user satisfaction data** from the annual Adult Social Care Survey to measure service users' ease of finding information and advice, satisfaction with care services, and perceptions about the impact of their care on quality of life. Finally, we also adopt two perceptual indicators from the biennial Personal Social Survey of Adult Carers measuring satisfaction of informal carers with council-run services.[13] Overall, these additional measures capture aspects of council performance that extend beyond the Ombudsman's narrow remit of maladministration whilst being potentially amenable to any improvement initiatives that managers may attempt after censure.

Turning to the expected behaviors during subunit turnarounds:

To capture leadership attention (H2), we record the **number of formal decisions issued each year** by each council about its adult social care services**.** Councils publish lists of decisions taken at cabinet meetings and other committees, typically designating these as "key" and "non-key" decisions, respectively. We obtained and coded every decision related to adult social care during our study period, creating a panel dataset to track committee attention toward this service over time. Additionally, to measure the degree of centralized control over adult social care departments (H3),

[13] As this survey is conducted biennially, we assume that the situation in non-survey years remains consistent with the previous survey year, to align with our annual panel analysis.

we obtained internal audit plans from council websites and recorded the **number of audit days assigned to adult social care each year**. As Power (1999, p. 6) explains, audit committees function as "internal agents of control," and internal audit plans are documents that specify where an organization's control risks lie, and how the internal audit staff is to be deployed to provide internal assurance against those risks. Thus, a higher allocation of audit time signals an increased concern to secure compliance in that subunit, consistent with O'Kane and Cunningham's advice to "instill close supervision" during turnarounds (already cited).

To observe the rate of investment in core competencies (H4), we measure **staff recruitment patterns** using the Adult Social Care Workforce Dataset, which is collated annually by central government from local statistical returns. This includes information on over 110,000 roles within 151 councils in categories including frontline, professional and supervisory roles. Frontline includes staff interacting directly with the public, other than regulated professionals; while supervisors manage and coordinate frontline and professional teams. In addition, we use data from Municipal Yearbooks to track **turnover in director-level** leadership of adult social care in each council.

Finally, because performance may be influenced by local demographic conditions, we control for the proportion of adults in the local population and the proportion with long-term health conditions, using Office for National Statistics data.

Variable definitions and data sources are provided in Appendix Tables A1 and A2.

4.2 Empirical strategy

Our study aims to provide a credible causal explanation of the impact of the Ombudsman's public interest notifications on local government performance and behaviors. The key challenge is ensuring that our treatment and control groups are fully comparable, given their non-random assignment. The Ombudsman can only notify councils about which it has both received and upheld a complaint from the public. (That is, it has no right to self-initiated investigations.)[14] In addition, escalation to the public interest report stage is largely a discretionary choice, albeit justified *internally* according to the aforementioned criteria. To ensure the accuracy of our Average Treatment Effect on the Treated estimates, we first employ matching to maximize comparability between treatment and control groups, and then apply staggered difference-in-differences to infer the effects of the Ombudsman's report releases over time.

We use the Coarsened Exact Matching (CEM) approach proposed by Iacus et al. (2012) and widely used in economics (see Aneja & Xu, 2021; Kotsogiannis et al., 2024). CEM is an exact matching algorithm that partitions

[14] If, during the course of an investigation, a potential further injustice is uncovered, LGSCO may also investigate it without receiving a further complaint (Thomas, 2024).

data into strata based on all possible combinations of predefined observable bins. Compared to standard matching methods, CEM enhances balance, reduces model dependence, and decreases estimation error. The matching variables are first coarsened (divided into discrete categories) and then exact matches are made based on these categories. This process yields counterfactuals that are comparable in terms of the joint distribution of observable baseline characteristics. (We recognize that the Ombudsman's discretionary decision-making may also reflect unobservable factors that cannot be accounted for in our data.) We use the Ombudsman's case management data to quantify the four aforementioned criteria for regulatory escalation: recurring faults, significant injustices, systemic problems, and repeated complaints about a single issue. To operationalize these criteria, we create seven indicators: the number of complaints received, investigated, and upheld; and the major categories of recommendations (apologies, financial redress, procedure reviews, and staff training). In addition, we include the proportion of the population with long-term health conditions and the proportion of adults in the population.

Table 1 shows that, before matching, there were several significant differences between the treated and control groups – in total complaints, investigated complaints, financial redress cases, and the rate of health problems in the population (Columns 1 and 2). Through matching, we found suitable counterparts for 48 out of 53 treated councils, and for 85 out of 98 control councils. Column 4 reports the mean differences between matched

treated and control councils in 2012. Notified and unnotified councils are now highly comparable, with no statistically significant differences in complaints, recommendations or demographic covariates.

In matching councils using 2012 data, one further concern is that these attributes may fluctuate significantly during the remainder of the study period, reducing the quality of the match. Therefore, in Figure A1 in the Appendix, we examine correlations between the values of these covariates in 2012 and their average values from 2013 to 2019. We find these variables to exhibit long-term stability (close to a 45-degree linear correlation), indicating that our approach is sufficient to capture the long-term characteristics of local councils. (We refrain from implementing a dynamic matching strategy because the staggered timing of treatment necessitates stable untreated controls over time, and re-matching would distort the control group composition and induce post-treatment bias.)

Table 1: Descriptive statistics of whole and matched samples in 2012

	(1) Whole sample mean	(2) Treated-control diff.	(3) Matched sample mean	(4) Matched treated-control diff.
Panel A: Ombudsman complaints and recommendations				
number of total complaints (logged)	1.0506	0.591***	0.9393	0.222
		(0.139)		(0.142)
number of investigated complaints (logged)	0.4609	0.309***	0.3767	-0.00561
		(0.110)		(0.105)
number of upheld complaints (logged)	0.1965	0.114	0.1404	-0.0261
		(0.0695)		(0.0658)
number of apologies (logged)	0.1236	0.0628	0.0749	0.0362
		(0.0525)		(0.0449)
number of financial redress cases (logged)	0.1511	0.112**	0.0976	0.0001
		(0.0543)		(0.0530)
number of procedure reviews (logged)	0.0677	0.0231	0.0103	-0.0001
		(0.0393)		(0.0167)
number of staff training (logged)	0.0275	0.0110	0.0154	0.0131
		(0.0234)		(0.0197)
Panel B: Population demographics				
rate of health problems in the population (%)	0.1319	0.00605**	0.1317	0.00353
		(0.00303)		(0.00328)
adult rate in the population (%)	0.8011	0.000435	0.8016	0.000121
		(0.00397)		(0.00429)
Sample size		151		133
of which treated upper-tier local authorities		53		48
of which control upper-tier local authorities		98		85

Note: The table compares the whole sample and matched local authorities in 2012. Column 1 shows mean characteristics for the whole sample, and Column 2 reports mean differences between treated and control groups. Columns 3 and 4 present the same metrics for the matched sample (using coarsened exact matching).

Using the trimmed sample, we estimate the dynamic effects of the Ombudsman's public interest notifications using a staggered difference-in-differences (DiD) event study framework (Sun & Abraham, 2021). This extends conventional DiD by accounting for staggered treatment and addressing biases from heterogeneous treatment timing, yielding robust causal estimates. It captures both immediate reactions and longer-term trends in

councils' behavior. We implement Callaway and Sant'Anna's (2021) group-time average treatment effect estimator to accommodate these features.

For council *i* in year *t*, let E_i be the period when council *i* is first treated. Then let $K_{it} = t - E_i$ be the number of years before or after the event. We regress outcome Y_{it} on 1 ($K_{it} = k$) relative year indicators for the fully saturated set of indicators going from the beginning to end of the sample:

$$Y_{it} = \alpha + \sum_k \beta_k [1(K_{it} = k)] + \gamma_i + \delta_t + \varepsilon_{it} \qquad (1)$$

In this specification, β_k is the Average Treatment Effect on the Treated (ATT) estimate at length of exposure *k* from the first report releasing. We control for council and year fixed effects, denoted by γ_i and δ_t respectively. We also control for demographic conditions. Standard errors are clustered at the council level.

5. RESULTS

Using the combined CEM and staggered DiD strategies, we begin by testing the main effects of low-stakes accountability on future council performance, and then turn to the possible mechanisms through which performance improvements occur.

5.1 Turnaround outcomes

Our main indicator of turnaround is whether the future incidence of maladministration reduces. Table 2, Panel A, Column 1 presents the baseline specification for Equation (1), with a basic set of fixed effects without additional controls. This indicates that, on average, councils significantly reduced upheld complaints following receipt of a public interest report. The estimates in Column 2, which add demographic controls, remain largely unchanged in both significance and magnitude — a 57% reduction in maladministration after notification, corresponding to approximately 3.05 fewer Ombudsman-upheld cases per council per year.[15] Columns 3 and 4 present a placebo test using volume of complaints *not* accepted for investigation as the outcome variable. This tests whether our findings reflect broader reporting trends or unrelated factors, rather than genuine

[15] Percentage change is calculated as (e^(-0.844)-1)*100=-57%. Given the mean of 5.347 upheld cases, the estimated treatment is 5.347*e^(-0.844)=2.301. This corresponds to an absolute reduction of 3.05 cases on average. The 99% confidence interval implies that the percentage reduction may range from approximately 4% to 81%, representing an absolute reduction between 0.21 and 4.32 cases on average.

performance gains. The analysis shows no significant change in unaccepted complaints after notification, reinforcing our interpretation of H1.

Table 2: Effects of low-stakes accountability on performance turnaround

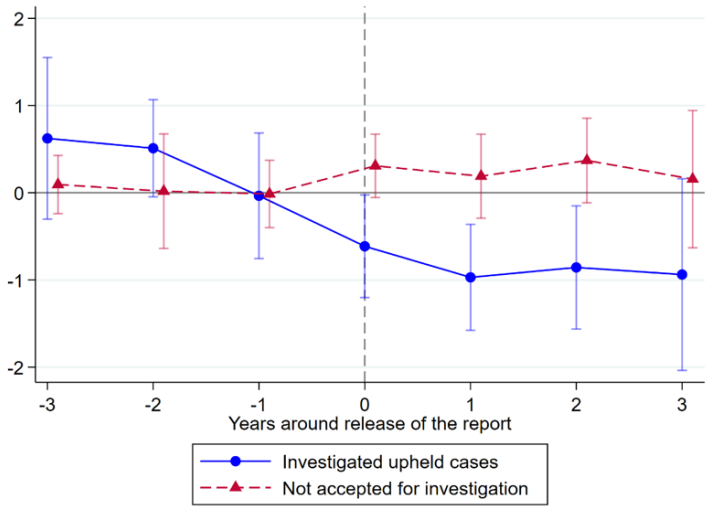
	(1)	(2)	(3)	(4)
	Investigated upheld cases		Not accepted for investigation	
Panel A				
Post report releasing	-0.820***	-0.844***	0.275	0.256
	(0.305)	(0.312)	(0.254)	(0.245)
Panel B				
Year 0	-0.637**	-0.614**	0.324*	0.308*
	(0.284)	(0.300)	(0.193)	(0.185)
Year 1	-0.907***	-0.970***	0.206	0.190
	(0.315)	(0.310)	(0.244)	(0.246)
Year 2	-0.839**	-0.856**	0.373	0.370
	(0.347)	(0.361)	(0.263)	(0.247)
Year 3	-0.898	-0.938*	0.198	0.156
	(0.549)	(0.561)	(0.407)	(0.401)
Local authority FEs	✓	✓	✓	✓
Year FEs	✓	✓	✓	✓
Demographic control		✓		✓
Empirical method	CEM+DiD	CEM+DiD	CEM+DiD	CEM+DiD
Observations	1,064	1,048	1,064	1,048

Notes: The unit of observation is the local authority × year. Post report releasing is a dummy that is 1 if local authority has received the report, and 0 otherwise. Standard errors clustered at the local authority level. *** p<0.01, ** p<0.05, * p<0.1.

The DiD design assumes that, absent the shock event (notification), treated and control councils would have experienced similar trends in outcome variables. Although this is not directly testable, we follow convention by testing its validity for pre-treatment trends. Figure 3 depicts the impact of notification on both upheld complaints and cases not accepted for investigation, using year-by-year estimates. Before receiving the Ombudsman's report (year –1 and earlier), there is no significant difference at the 5% level between treated and control councils in the number of upheld complaints, indicating parallel trends. However, after notification,

treated councils show a clear reduction in upheld cases relative to the matched control group. This trend is corroborated by the year-by-year estimates presented in Table 2, Panel B. The effect becomes apparent in the year of notification (year 0), with an approximate 47.1% decrease, reaching a peak in the following year (year 1) with a decline of around 59.6%. Conversely, in the placebo test, the number of cases *not* accepted for investigation remains stable before and after notification, again suggesting that the observed improvement in maladministration is not being driven by general changes in complaint reporting.

Figure 3: Event study - Performance turnaround in censured vs. uncensured councils



A further concern is that our results might be driven entirely by the influence of influence of never-treated councils in the staggered design, rather than reflecting the temporal differences between councils receiving reports. We therefore also adopt the most conservative strategy by conducting staggered DiD analysis using only councils that received public interest

reports during the study period, excluding the never-treated group. Even with this smaller sample, our baseline findings remain supported (see Table A5 and Figure A2 in the appendix).

Finally, as noted, several councils receive two or even three public interest reports related to adult social care during our study period. So far, we only accounted for the first moment of notification. In the appendix, we modify our approach to explore the impact of subsequent notifications. As explained there, these additional analyses all suggest that the performance improvement is driven largely by the initial notification, with little evidence of incremental effects from subsequent reports.

With the main effect established, we next investigate whether performance improvements are confined narrowly to the Ombudsman's remit or are more broad-based (H1a). In Table 3, we use a combination of administrative and survey data to assess council performance beyond maladministration. Across the eight columns in Panel A of Table 3, we find that censuring by the Ombudsman induced no significant effect on council-attributable delayed discharges from hospital, or satisfaction among either care users or informal carers. In Figure 4 and Panel B of Table 3, we further examine year-by-year estimates. The coefficients for each year following notification show that this maladministration-focused accountability regime *did not* have significant effects on performance beyond the account-holder's remit, consistent with H1a.

Figure 4: Event study - Tunnel vision in censured vs. uncensured councils

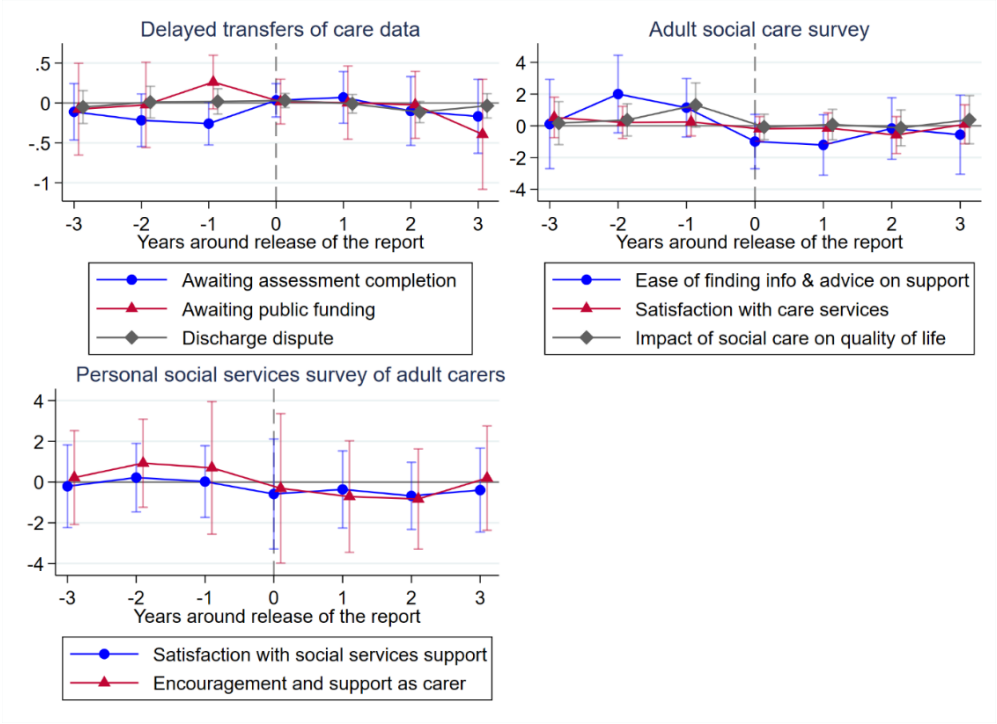


Table 3: Effects of low-stakes accountability on performance beyond maladministration

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Delayed Transfers of Care Data			Adult Social Care Survey			Personal Social Services Survey of Adult Carers	
	Awaiting assessment completion	Awaiting public funding	Discharge dispute	Ease of finding info & advice on support	Satisfaction with care services	Impact of social care on quality of life	Encouragement and support as carer	Satisfaction with social services support
Panel A								
Post report releasing	-0.041	-0.098	-0.031	-0.732	-0.205	0.064	-0.506	-0.415
	(0.158)	(0.178)	(0.041)	(0.802)	(0.430)	(0.470)	(0.920)	(1.183)
Panel B								
Year 0	0.034	0.018	0.031	-0.990	-0.192	-0.072	-0.585	-0.309
	(0.107)	(0.143)	(0.046)	(0.874)	(0.395)	(0.410)	(1.380)	(1.873)
Year 1	0.070	0.004	-0.010	-1.208	-0.142	0.077	-0.362	-0.714
	(0.165)	(0.234)	(0.059)	(0.975)	(0.488)	(0.490)	(0.967)	(1.397)
Year 2	-0.100	-0.023	-0.113*	-0.171	-0.580	-0.130	-0.679	-0.832
	(0.220)	(0.214)	(0.068)	(0.989)	(0.593)	(0.576)	(0.842)	(1.255)
Year 3	-0.167	-0.392	-0.034	-0.560	0.093	0.381	-0.395	0.195
	(0.236)	(0.351)	(0.078)	(1.270)	(0.627)	(0.775)	(1.052)	(1.307)
Local authority FEs	✓	✓	✓	✓	✓	✓	✓	✓
Year FEs	✓	✓	✓	✓	✓	✓	✓	✓
Demographic control	✓	✓	✓	✓	✓	✓	✓	✓
Empirical method	CEM+DiD	CEM+DiD	CEM+DiD	CEM+DiD	CEM+DiD	CEM+DiD	CEM+DiD	CEM+DiD
Observations	1,040	1,040	1,048	1,026	1,016	1,018	1,048	1,048

Notes: The unit of observation is the local authority × year. Post report releasing is a dummy that is 1 if local authority has received the report, and 0 otherwise. Post is a dummy that is 1 if the year is after the report releasing. Standard errors clustered at the local authority level. *** p<0.01, ** p<0.05, * p<0.1.

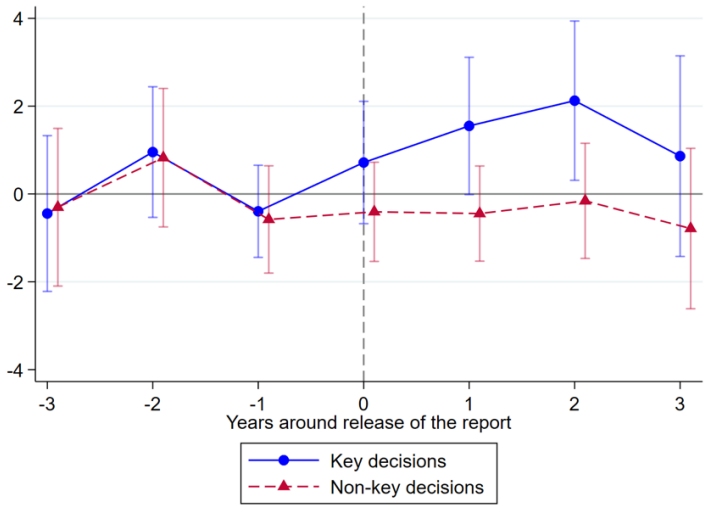
5.2 Leadership attention

Having confirmed the main effect of notification on turnaround success (narrowly conceived), we next explore the hypothesized behaviors.

To measure any changes in leadership attention, we quantify the number of key decisions about adult social care taken by council cabinets. If attention is redirected by the low-stakes accountability event, we expect a noticeable increase in top-level decision-making related to adult social care. Figure 4 presents an event study where the solid blue line depicts the yearly impact of notification on the number of key decisions. Prior to treatment, the coefficients for key decisions hover around zero, indicating no significant differences between notified and unnotified councils. However, following censure, decision-making on adult social care increases, peaking in the second year with around 9.08 additional top-level decisions on average, relative to the pre-censure period. This suggests that leadership attention is indeed significantly redirected after notification from the Ombudsman is received, supporting H2.[16]

[16] In the appendix, Table A6 Column 1 provides the specific coefficient estimates illustrating this effect.

Figure 5: Event study – Leadership attention in censured vs. uncensured councils



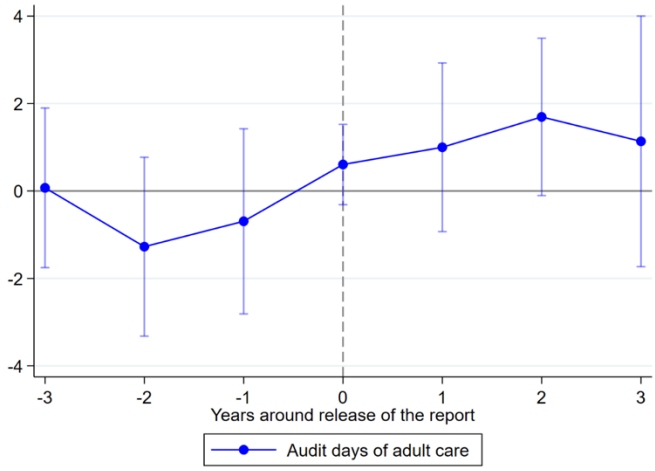
To validate this finding, we again conduct a placebo test, now using "non-key" decisions by more junior and specialist committees as the outcome variable. If the observed increase in key decisions is truly attributable to heightened *leadership* attention, we should not see a similar increase in non-key decisions. The red dashed line in Figure 5 and the results in Table A6 in the appendix confirm this expectation. There is no significant change in non-key decision-making related to adult care following censure.

5.3 Centralization

Next, we investigate whether notification leads to centralization of control over the responsible subunit, using our dataset of internal audit plans. Figure 6 shows the annual change in audit days allocated to adult care following the public interest notification. We observe a gradual increase in the year-by-year point estimates, with a significant rise in audit days in the second year after notification. This suggests that councils are deploying control resources

to supervise adult social care departments, consistent with H3, although this effect is significant only at the 10% level – likely due to missing data.[17]

Figure 6: Event study – centralized control in censured vs. uncensured councils



5.4 Urgent investments in core competencies

Investing in core competencies is a further means of performance turnaround (H4). Following Boyne and Meier (2009), we operationalize this as changes to workforce composition. Table 4, Column 1 shows that censured councils grow their adult social care workforce considerably, compared with the control group. Columns 2 to 4, and Figure 8, present event study results by job groupings, revealing a particularly pronounced increase in frontline roles. Specifically, Panel B of Table 4, Column 2 indicates that these roles increased by in the order of 17.58 posts in the first year after censure, rising to 30.47 in

[17] Only 61% of council-years have internal audit plans available on a comparable and code-able basis. In the appendix, Table A7 provides the specific coefficient estimates.

the second. By contrast, Columns 3 and 4 of Table 4 show no significant increases in supervisory or professional roles, with coefficients that are neither economically nor statistically significant at any point. Together, this indicates a targeted investment in public-facing (rather than professional or back-office) roles, consistent with turnaround theories predicting a focus on core business.

Table 4: Effects of low-stakes accountability on core competencies

	(1)	(2)	(3)	(4)
	Overall	Job categories		
	workforce	frontline	supervisory	professional
Panel A				
Post report releasing	0.032	0.136**	-0.100	-0.105
	(0.059)	(0.063)	(0.082)	(0.077)
Panel B				
Year 0	-0.012	-0.006	-0.069	-0.017
	(0.038)	(0.046)	(0.067)	(0.048)
Year 1	0.030	0.134**	-0.091	-0.132*
	(0.054)	(0.060)	(0.084)	(0.079)
Year 2	0.094	0.222***	-0.096	-0.099
	(0.074)	(0.079)	(0.093)	(0.101)
Year 3	0.017	0.194	-0.146	-0.171
	(0.104)	(0.120)	(0.145)	(0.129)
Local authority FEs	✓	✓	✓	✓
Year FEs	✓	✓	✓	✓
Demographic control	✓	✓	✓	✓
Empirical method	CEM+DiD	CEM+DiD	CEM+DiD	CEM+DiD
Observations	1,010	998	1,009	1,003

Notes: The unit of observation is the local authority × year. Post report releasing is a dummy that is 1 if local authority has received the report, and 0 otherwise. Post is a dummy that is 1 if the year is after the report releasing. Standard errors clustered at the local authority level. *** p<0.01, ** p<0.05, * p<0.1.

We also examined whether notification influenced turnover in the director position responsible for adult social care within councils' executive teams. Figure 9 shows no discernible effect of censure on turnover, with a flat trend in the three years before and after the report was issued. As already indicated,

this is consistent with Ombudsmen providing a low-stakes accountability regime, without threat to job security. But it also demonstrates that the improvement in performance that we observe are not explained by leadership "takeover."

Figure 8: Event study – Core competencies in censured vs. uncensured councils

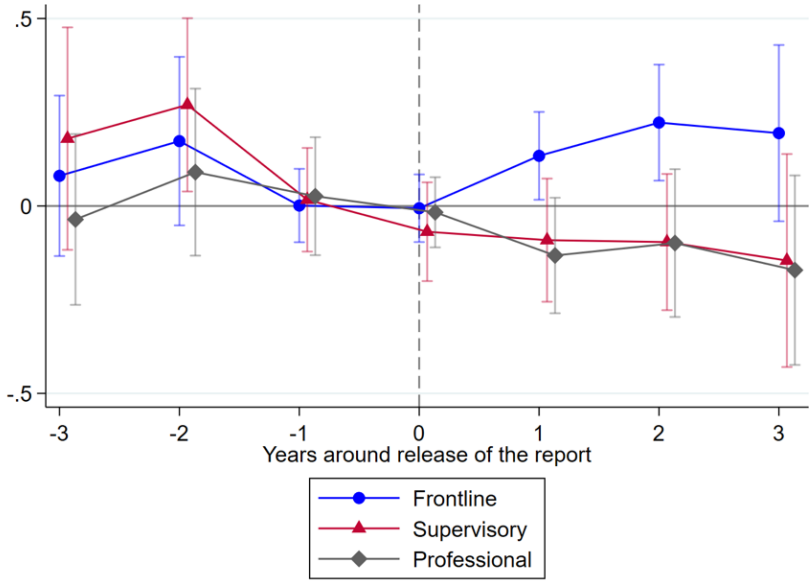
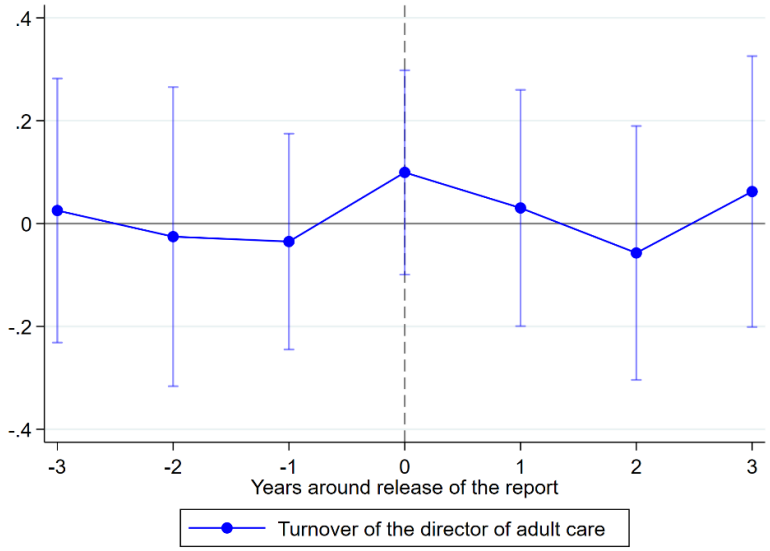


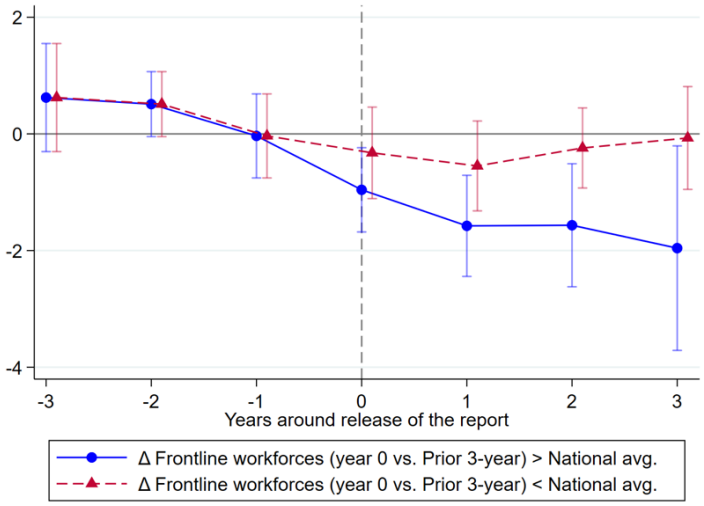
Figure 8: Event study – Turnover of director in censured vs. uncensured councils



Finally, we examine whether *more rapid* investment in core competencies leads to greater performance improvements (H5). We compare the number of frontline roles in the censure year (year 0) with the average in the three preceding years. This comparison yields the immediate change rate in hiring during the year of censure. We then classify organizations whose change rate exceeds the national average into the "rapid-response" group, and those below this into the "delayed-response" group, and compare their performance in reducing maladministration.[18] The results, in Figure 9, indicate that prompt investments deliver significantly greater performance recovery, consistent with the turnaround adage that "time is of the essence" (H5). In the first year, the rapid-response group saw a more substantial reduction in upheld complaints compared to the delayed-response group. This trend continues over the next three years, with the faster-responding group demonstrating consistently greater decreases in upheld cases. As such, prompt re-allocation of resources appears to accelerate the turnaround process, leading to more significant and sustained performance gains over time.

[18] In the appendix, Table A9 provides the specific coefficient estimates illustrating this effect.

Figure 9: Event study - Effects of urgency on maladministration



6. DISCUSSION

6.1. Accountability-induced turnarounds

We set out to establish whether accountability can induce turnaround in struggling organizations (or parts thereof), and, further, to see if feedback and standard setting alone can provide sufficient impetus for change. Our results provide consistent evidence that low-stakes accountability, without high-powered incentives, is able to trigger performance recovery in organizational subunits. This is consistent with recent theoretical work (Meier et al., 2015; Woo et al., 2015; Jakobsen et al., 2018; Andersen & Nielsen, 2020), and with legal analysis on the surprising degree of influence obtained by low-powered ombudsmen, including Hertogh's (2001) seminal comparison with administrative courts in the Netherlands. (Our analysis does not, of course, reveal whether high- or low-stakes accountability performs best in a head-to-head comparison. That will be for future research to judge in various scenarios.)[19]

Nevertheless, we were unable to detect any broader performance gains beyond maladministration. This reflects our expectation of some residual tunnel vision, even in low-stakes accountability regimes. Measurement

[19] At present, there are few direct comparisons of high- and low-stakes regimes. One example is Hanushek and Raymond's (2005) analysis of different states' enactment of No Child Left Behind legislation, in which low-stakes "report cards" are found to achieve minimal gains compared with more "consequential" systems involving monetary awards or takeover threats.

encourages effort prioritization; and narrow feedback from specialist accountability bodies fosters restrictive rather than broad-based learning. Accountability-induced turnarounds thus appear to be *partial* turnarounds, limited by the remit and/or priorities of the account-holder. As Bovens et al. (2008, p. 228) write, "administrators … get better in meeting the requirements posed by their accountability forums – but not necessarily … in the real world of … public service delivery." This suggests that accountability needs to be complemented by other turnaround strategies if it is to improve extra-jurisdictional dimensions.

Besides these main effects, our multidimensional dataset also affords valuable opportunities to test key turnaround behaviors. Our evidence confirms that council leaders significantly increase decision-making in respect of adult social care following the Ombudsman's notifications, consistent with the idea that turnaround needs a "triggering" event. Turnaround also typically involves exercising closer control over problem areas (O'Kane & Cunningham, 2014; Meier et al., 2015), and switching investments from ancillary to core activities (McKiernan, 2006; Boyne & Meier, 2009; Murphy, 2010). Again, we found substantially greater internal audit days allocated toward the responsible subunit following notification, and a large increase in frontline staffing. Moreover, when these investments are implemented rapidly, the long-term improvement in performance is materially greater, consistent with several empirical studies of private sector recoveries (Tangpong et al., 2015; Barbero et al., 2017).

All that said, the behavioral changes we studied mostly appear to accompany rather than proceed the main turnaround outcome. Maladministration is significantly reduced *within* the same year as the public interest notification is received, and the improvement peaks in the follow year. In contrast, measurable changes in leadership attention, centralization, and staff recruitment only emerge a year *after* the notification is given, and do not reach their respective peaks until two years later. This not only indicates that the "Ombudsman effect" is long lasting (and hardly that of an Ombuds*mouse*), but also that the mechanisms of turnaround may be more indirect than previously supposed – something we address in following subsection.

6.2 Theoretical implications

Based on these results, we are left, firstly, with the question of why the low-powered Ombudsman is able to induce such a strong reaction. There are several possibilities. The LGSCO has direct access to board-level decision-makers in councils, guaranteed in statute. Many of its investigators are "poachers turned game keepers," drawn from the ranks of local government and so knowledgeable about the sector and its challenges. The non-coercive and low-stakes environment may encourage dialogue and reflection, rather than blame avoidance and self-preservation (Hertogh, 2001; Gill, 2018). But perhaps most notably, the LGSCO's feedback to councils is entirely based on the lived experience of individuals who depend upon the organization's services and have been somehow "let down" by it. As Thomas

(2024) writes, "the LGSCO is a casework institution. It sees and approaches local government through individuals' complaints and how councils' failings cause people injustice." Given the dilemma we outlined previously – that the more belief-inconsistent the account-holder's feedback is, the more *credible* it must be – the rooting of the Ombudsman's reports in real, personal injustices may offer a significant advantage when it comes to revealing the "gap" between aspirations and achievements necessary to trigger remedial action (Meier et al., 2015; Andersen & Nielsen, 2020). The potential galvanizing effects of this combination of lived experience and moral authority is a promising line of research.

Alternatively, while ombudsmen fail to meet Hibbard et al.'s (2003) four criteria for inflicting external reputational damage on account-givers, *internal* leader reputation among peers may also partly explain our results. Top council officials each oversee subunits that are subject to the Ombudsman's jurisdiction (housing, highways, etc.); each know what a public interest notification signifies; and, by virtue of their seniority, are each aware every time a report is issued rebuking one of their peer's departments. To the extent that this causes embarrassment among colleagues, action may be taken to "save face." These *internal* reputational effects should be explored in future research.

Finally, while we observe the whole series of changes in performance and behavior predicted by our theory, the temporal ordering is puzzling. Maladministration improves rapidly, whereas organizational changes are

slower and more sustained. Our tentative interpretation is that, once councils are notified of failures, staff in responsible subunits proactively seek to recover performance *in anticipation* of the increase in leaders' attention and control that will now be forthcoming. This is consistent with recent literature on felt accountability (Schillemans et al., 2021), with the Ombudsman's notification materially increasing the subunit's *expectation* of being called to account by council leaders. This departs from the conventional understanding in turnaround research that attention and centralization are direct mechanisms by which recovery is achieved, and so requires further quantitative and qualitative investigation.

6.3 Limitations

The principal limitation to our study is the extent to which the control group, matched with the treated group on observable characteristics at baseline, provides a convincing counterfactual throughout our study period. There may be unobserved factors that led to the Ombudsman selecting a council for a public interest notification that are not incorporated into our matching strategy. Moreover, some councils that narrowly avoided the notification treatment early in the period may have self-initiated improvements in their adult care services, rendering themselves less comparable to later-treated councils. To address these challenges, we perform a suite of robustness tests to ascertain the stability of our results. We demonstrate the correlation between our matching covariates in 2012 and their averages from 2013 to 2019. And we focus exclusively on councils that received reports during the

study period, thereby isolating the effect of the timing of the event's impact. These robustness checks demonstrate that our baseline empirical strategy, combining CEM and DiD, is effective.

7. CONCLUSION

We theorized that low-stakes accountability would produce turnaround in under-performing local government services, driven by increased leadership attention to and control over problem areas, and investment in core competencies. Using panel data, coarsened exact matching, and staggered difference-in-differences, we provided strong evidence in support of these hypotheses, whilst also confirming the partial nature of accountability-induced turnarounds. Our study adds to the few existing quantitative evaluations of public service turnaround strategies, including Boyne and Meier's (2009), Rutherford (2014), and Alonso & Andrews's (2021), which is also quasi-experimental. It also confirms the presence of key turnaround behaviors observed in the corporate sector – like centralization and urgent investments – in the government context, whilst also suggesting that some recovery occurs in anticipation of increased leader attention, rather than in consequence of it. And finally, our results challenge the consensus in political science that, for want of strong enforcement mechanisms, the public sector is unable to regulate *itself* (Wilson & Rachal, 1977; Konisky & Teodoro, 2016). Indeed, in spite of – or perhaps because of – a lack of enforcement powers, the ombudsman is a remarkably influential regulator "inside" government (Hood et al., 1999).

REFERENCES

Aleksovska, M., Schillemans, T., & Grimmelikhuijsen, S. (2019). Lessons from five decades of experimental and behavioral research on accountability: A systematic literature review. *Journal of Behavioral Public Administration*, *2*(2), 1-18.

Alonso, J. M., & Andrews, R. (2021). Government-Created Nonprofit Organizations and Public Service Turnaround: Evidence from a Synthetic Control Approach. *Journal of Public Administration Research and Theory*, *31*(2), 346-362.

Andersen, L. B., Boesen, A., & Pedersen, L. H. (2016). Performance in Public Organizations: Clarifying the Conceptual Space. *Public Administration Review*, *76*(6), 852-862.

Andersen, S. C., & Nielsen, H. S. (2020). Learning from Performance Information. *Journal of Public Administration Research and Theory*, *30*(3), 415-431.

Aneja, A., & Xu, G. (2021). The Costs of Employment Segregation: Evidence from the Federal Government Under Woodrow Wilson. *The Quarterly Journal of Economics*, *137*(2), 911-958.

Argyris, C. (1977). Double loop learning in organizations. *Harvard Business Review*, *55*(5), 115.

Ayres, I., & Braithwaite, J. (1992). *Responsive Regulation: Transcending the Deregulation Debate*. Oxford University Press.

Barbero, J. L., Di Pietro, F., & Chiang, C. (2017). A rush of blood to the head: Temporal dimensions of retrenchment, environment and turnaround performance. *Long Range Planning*, *50*(6), 862-879.

Bevan, G. (2010). Performance Measurement of "Knights" and "Knaves": Differences in Approaches and Impacts in British Countries after Devolution. *Journal of Comparative Policy Analysis*, *12*(1-2), 33-56.

Bevan, G. (2021). *Delivering transformational change for England's National Health Service*. Retrieved 22nd July from https://learninggeneration.org/the-uk-prime-ministers-delivery-unit-pmdu-delivering-transformational-change-for-englands-national-health-service/

Bevan, G., & Hood, C. (2006). What's measured is what matters: targets and gaming in the English public healthcare system. *Public Administration*, *84*(3), 517-538.

Bovens, M., Schillemans, T., & Hart, P. T. (2008). Does public accountability work? An assessment tool *Public Administration*, *86*(1), 225-242.

Bovens, M., & Wille, A. (2021). Indexing watchdog accountability powers a framework for assessing the accountability capacity of independent oversight institutions. *Regulation & Governance*, *15*(3), 856-876.

Boyne, G. A. (2006). Strategies for Public Service Turnaround. *Administration & Society*, *38*(3), 365-388.

Boyne, G. A. (2010). Performance management: does it work? In R. M. Walker, G. A. Boyne, & G. A. Brewer (Eds.), *Public Management and Performance: Research Directions* (pp. 207-226). Cambridge University Press.

Boyne, G. A., & Meier, K. J. (2009). Environmental Change, Human Resources and Organizational Turnaround. *Journal of Management Studies*, *46*(5), 835-863.

Brewer, G. A. (2006). All measures of performance are subjective: More evidence on US federal agencies. In G. A. Boyne, K. J. Meier, J. L. J. O'Toole, & R. M. Walker (Eds.), *Public Service Performance: Perspectives on Measurement and Management* (pp. 35-54). Cambridge University Press.

Burgess, S., Wilson, D., & Worth, J. (2013). A natural experiment in school accountability: The impact of school performance information on pupil progress. *Journal of Public Economics*, *106*, 57-67.

Burn, E., Redgate, S., Needham, C., & Peckham, S. (2024). Implementing England's Care Act 2014: was the Act a success and when will we know? *International Journal of Care and Caring*, *8*(1), 47-63.

Caillier, J. G. (2023). A Public Organization Provided a Poor Service: Is There Anything They Can Do to Make It Right with the Citizen? *Public Performance & Management Review*, *46*(6), 1469-1501.

Campbell, D. T. (1979). Assessing the impact of planned social change. *Evaluation and Program Planning*, *2*(1), 67-90.

Castelló-Sirvent, F., & Roger-Monzó, V. (2023). Research agenda on turnaround strategies beyond systemic disruptions. *Journal of Organizational Change Management*, *36*(1), 86-105.

Clark, T. (2021). From The Silent Watchdog to the Lost Watchdog. *Media History*, *27*(2), 177-196.

de Wolf, I. F., & Janssens, F. J. G. (2007). Effects and side effects of inspections and accountability in education: an overview of empirical studies. *Oxford Review of Education*, *33*(3), 379-396.

Dee, T. S., & Jacob, B. (2011). The impact of No Child Left Behind on student achievement. *Journal of Policy Analysis and Management*, *30*(3), 418-446.

Deming, D. J., Cohodes, S., Jennings, J., & Jencks, C. (2016). School Accountability, Postsecondary Attainment, and Earnings. *The Review of Economics and Statistics*, *98*(5), 848-862.

Döring, M. (2022). 'Explain, but make no excuses': service recovery after public service failures. *Public Management Review*, *24*(5), 799-817.

Dubnick, M. (2005). Accountability and the promise of performance. *Public Performance & Management Review*, *28*(3), 376-417.

Dunleavy, P., Bastow, S., Tinkler, J., Goldchluk, S., & Towers, E. (2010). Joining up citizen redress in UK central government. In M. Adler (Ed.), *Administrative Justice in Context* (pp. 421-456). Hart.

Elston, T., Bel, G., & Wang, H. (2025). The effect of inter-municipal cooperation on social assistance programs: evidence from housing allowances in England. *Journal of Policy Analysis and Management*, *44*(3), 1060-1088.

Entwistle, T., & Doering, H. (2024). Amoral Management and the Normalisation of Deviance: The Case of Stafford Hospital. *Journal of Business Ethics*, *190*(3), 723-738.

Gerrish, E. (2016). The Impact of Performance Management on Performance in Public Organizations: A Meta-Analysis. *Public Administration Review*, *76*(1), 48-66.

Gill, C. (2018). What can government learn from the ombudsman? In M. Hertogh & R. Kirkham (Eds.), *Research Handbook on the Ombudsman*. Edward Elgar.

Glasby, J. O. N., Zhang, Y., Bennett, M. R., & Hall, P. (2021). A lost decade? A renewed case for adult social care reform in England. *Journal of Social Policy*, *50*(2), 406-437.

Goodair, B., McManus, A., Esposti, M. D., & Bach-Mortensen, A. (2024). How outsourcing has contributed to England's social care crisis. *BMJ*, *387*.

Goodhart, C. A. (1981). Problems of monetary management: the UK experience. In A. S. Courakis (Ed.), *Inflation, Depression and Economic Policy in the West* (pp. 111-144). Barnes & Noble.

Gopinath, C. (2005). Recognizing Decline: The Role of Triggers. *American Journal of Business*, *20*(1), 21-27.

Gulland, J. (2010). Current developments in the UK—complaints procedures and ombudsmen. In M. Adler (Ed.), *Administrative Justice in Context* (pp. 457-481).

Gulland, J. (2024). Administrative justice and austerity: the case of disputes about adult social care. In S. Cowan & S. Halliday (Eds.), *Essays in honour of Michael Adler* (pp. 187-207). Springer Nature.

Gwyn, W. B. (1973). The British PCA: "Ombudsman or ombudsmouse?". *The Journal of Politics*, *35*(1), 45-69.

Hanushek, E. A., & Raymond, M. E. (2005). Does school accountability lead to improved student performance? *Journal of Policy Analysis and Management*, *24*(2), 297-327.

Hargrove, E. C., & Glidewell, J. C. (Eds.). (1990). *Impossible Jobs in Public Management*. University of Kansas Press.

Harlow, C. (2018). Ombudsmen: 'hunting lions' or 'swatting flies'. In M. Hertogh & R. Kirkham (Eds.), *Research Handbook on the Ombudsman*. Edward Elgar.

Harlow, C., & Rawlings, R. (2021). *Law and Administration*. Cambridge University Press.

Harvey, G., Jas, P., & Walshe, K. (2015). Analysing organisational context: case studies on the contribution of absorptive capacity theory to understanding inter-organisational variation in performance improvement. *BMJ Quality & Safety*, *24*(1), 48.

Harvey, G., Skelcher, C., Spencer, E., Jas, P., & Walshe, K. (2010). Absorptive capacity in a non-market environment: A knowledge-based approach to analysing the performance of sector organizations. *Public Management Review*, *12*(1), 77-97.

Heinrich, C. J., & Marschke, G. (2010). Incentives and their dynamics in public sector performance management systems. *Journal of Policy Analysis and Management*, *29*(1), 183-208.

Hertogh, M. (2001). Coercion, Cooperation, and Control: Understanding the Policy Impact of Administrative Courts and the Ombudsman in the Netherlands. *Law & Policy*, *23*(1), 47-67.

Hibbard, J. H., Stockard, J., & Tusler, M. (2003). Does publicizing hospital performance stimulate quality improvement efforts? *Health Affairs*, *22*(2), 84-94.

Hodges, C. (2018). The private sector ombudsman. In M. Hertogh & R. Kirkham (Eds.), *Research Handbook on the Ombudsman* (pp. 53-71). Edward Elgard.

Hofer, S. I., Holzberger, D., & Reiss, K. (2020). Evaluating school inspection effectiveness: A systematic research synthesis on 30 years of international research. *Studies in Educational Evaluation*, *65*, 100864.

Hood, C., James, O., Jones, G., Scott, C., & Travers, T. (1999). *Regulation Inside Government: Waste-Watchers, Quality Police, and Sleaze-Busters*. Oxford: Oxford University Press.

Iacus, S. M., King, G., & Porro, G. (2012). Causal Inference without Balance Checking: Coarsened Exact Matching. *Political Analysis*, *20*(1), 1-24.

Jakobsen, M. L., Baekgaard, M., Moynihan, D. P., & van Loon, N. (2018). Making Sense of Performance Regimes: Rebalancing External Accountability and Internal Learning. *Perspectives on Public Management and Governance*, *1*(2), 127-141.

Jas, P., & Skelcher, C. (2005). Performance decline and turnaround in public organizations. *British Journal of Management*, *16*(3), 195-210.

Kelman, S., & Friedman, J. N. (2009). Performance improvement and performance dysfunction: an empirical examination of distortionary impacts of the emergency room wait-time target in the English national health service. *Journal of Public Administration Research & Theory*, *19*(4), 917-946.

Kirkham, B. R. (2005). A Complainant's View of the Local Government Ombudsman. *Journal of Social Welfare and Family Law*, *27*(3-4), 383-394.

Kirkham, R., Thompson, B., & Trevor, B. (2008). When putting things right goes wrong: enforcing the recommendations of the ombudsman. *Public Law*, 510-530.

Konisky, D. M., & Teodoro, M. P. (2016). When Governments Regulate Governments. *American Journal of Political Science*, *60*(3), 559-574.

Kotsogiannis, C., Salvadori, L., Karangwa, J., & Mukamana, T. (2024). Do tax audits have a dynamic impact? Evidence from corporate income tax administrative data. *Journal of Development Economics*, *170*, 103292.

Le Grand, J. (2010). Knights and Knaves Return: Public Service Motivation and the Delivery of Public Services. *International Public Management Journal*, *13*(1), 56-71.

Levatino, A., Parcerisa, L., & Verger, A. (2024). Understanding the Stakes: The Influence of Accountability Policy Options on Teachers' Responses. *Educational Policy*, *38*(1), 31-60.

Local Government and Social Care Ombudsman (LGSCO). (2024a). *Annual Review of Adult Social Care Complaints, 2023-24*. Coventry: LGSCO

Local Government and Social Care Ombudsman (LGSCO). (2024b). *Manual for Councils*. Retrieved August 3 from https://www.lgo.org.uk/information-centre/information-for-organisations-we-investigate/councils/information-for-link-officers

Lynch, S., Mills, B., Theobald, K., & Worth, J. (2017). *Keeping your Head: NFER Analysis of Headteacher Retention*. National Foundaiton for Educational Research.

Mannion, R., & Braithwaite, J. (2012). Unintended consequences of performance measurement in healthcare: 20 salutary lessons from the English National Health Service. *Intern Med J*, *42*(5), 569-574.

McBurnie, G., Williams, J., & Coulter-Smith, M. (2023). The response from Scottish health boards to complaint investigations by the Scottish Public Services Ombudsman. *International Journal of Law in Context*, *19*(3), 407-426.

McKiernan, P. (2006). Turnarounds. In A. Campbell & D. O. Faulkner (Eds.), *Oxford Handbook of Strategy* (pp. 759-810). Oxford University Press.

Meier, K. J., Favero, N., & Zhu, L. (2015). Performance Gaps and Managerial Decisions: A Bayesian Decision Theory of Managerial Action. *Journal of Public Administration Research and Theory*, *25*(4), 1221-1246.

Meyers, C. V., & Murphy, J. (2007). Turning around Failing Schools. *Journal of School Leadership*, *17*(5), 631-659.

Moynihan, D. P. (2005). Goal-Based Learning and the Future of Performance Management. *Public Administration Review*, *65*(2), 203-216.

Murphy, J. (2010). Turning around failing organizations: Insights for educational leaders. *Journal of Educational Change*, *11*(2), 157-176.

Murphy, P., & Jones, M. (2016). Building the next model for intervention and turnaround in poorly performing local authorities in England. *Local Government Studies*, *42*(5), 698-716.

Needham, C., & Hall, P. (2023). *Social Care in the UK's Four Nations*. Policy Press.

O'Kane, C., & Cunningham, J. (2014). Turnaround leadership core tensions during the company turnaround process. *European Management Journal*, *32*(6), 963-980.

Parpworth, N. (2022). *Constitutional and administrative law*. Oxford University Press.

Paton, R., & Mordaunt, J. (2004). What's Different About Public and Non-Profit 'Turnaround'? *Public Money & Management*, *24*(4), 209-216.

Pearce, J. A., & Robbins, K. (1993). Toward Improved Theory and Research on Business Turnaround. *Journal of Management*, *19*(3), 613-636.

Perryman, J., Bradbury, A., Calvert, G., & Kilian, K. (forthcoming). Headteachers and external inspection in England. *Educational Management Administration & Leadership*.

Power, M. (1999). *The Audit Implosion: Regulating Risk from the Inside*. Centre for Business Performance, Institute of Chartered Accountants in England and Wales.

Propper, C., Sutton, M., Whitnall, C., & Windmeijer, F. (2010). Incentives and targets in hospital care: Evidence from a natural experiment. *Journal of Public Economics*, *94*(3), 318-335.

Romzek, B. S., & Dubnick, M. J. (1987). Accountability in the Public Sector: Lessons from the Challenger Tragedy. *Public Administration Review*, *47*(3), 227-238.

Rutherford, A. (2014). Organizational Turnaround and Educational Performance:The Impact of Performance-Based Monitoring Analysis Systems. *The American Review of Public Administration*, *44*(4), 440-458.

Rutherford, A., & Favero, N. (2020). Organizational turnaround: moving beyond one-size-fits-all solutions. *International Public Management Journal*, *23*(3), 315-335.

Schillemans, T. (2016). Calibrating Public Sector Accountability: Translating experimental findings to public sector accountability. *Public Management Review*, *18*(9), 1400-1420.

Schillemans, T., Overman, S., Fawcett, P., Flinders, M., Fredriksson, M., Laegreid, P., Maggetti, M., Papadopoulos, Y., Rubecksen, K., Rykkja, L. H., Salomonsen, H. H., Smullen, A., & Wood, M. (2021). Understanding Felt Accountability. *Governance*, *34*(3), 893-916.

Schweizer, L., & Nienhaus, A. (2017). Corporate distress and turnaround: integrating the literature and directing future research. *Business Research*, *10*(1), 3-47.

Seneviratne, M. (2008). Updating the local government ombudsman. *Public Law*, *4*, 627-635.

Simon, H. A. (1997). *Administrative Behavior*. Free Press.

Smith, P. (1995). On the unintended consequences of publishing performance data in the public sector. *International Journal of Public Administration*, *18*(2-3), 277-310.

Stacey, F. A. (1971). *The British Ombudsman*. Clarendon Press.

Staw, B. M., Sandelands, L. E., & Dutton, J. E. (1981). Threat Rigidity Effects in Organizational Behavior. *Administrative Science Quarterly*, *26*(4), 501-524.

Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, *225*(2), 175-199.

Tangpong, C., Abebe, M., & Li, Z. (2015). A Temporal Approach to Retrenchment and Successful Turnaround in Declining Firms. *Journal of Management Studies*, *52*(5), 647-677.

Thomas, R. (2024). The English Local Government and Social Care Ombudsman: Systemic Administrative Justice and Bureaucratisation Parts 1 and 2. *Public Law*, *2*, 424-447.

Thomassen, J. P., Leliveld, M. C., Van de Walle, S., & Ahaus, K. (2017). Compensating citizens for poor service delivery: Experimental research in public and private settings. *Public Administration*, *95*(4), 895-911.

Trahms, C. A., Ndofor, H. A., & Sirmon, D. G. (2013). Organizational Decline and Turnaround:A Review and Agenda for Future Research. *Journal of Management*, *39*(5), 1277-1307.

Tuggle, C. S., Sirmon, D. G., Reutzel, C. R., & Bierman, L. (2010). Commanding board of director attention: investigating how organizational performance and CEO duality affect board members' attention to monitoring. *Strategic Management Journal*, *31*(9), 946-968.

Van de Walle, S. (2016). When public services fail: a research agenda on public service failure. *Journal of Service Management*, *27*(5), 831-846.

Verger, A., & Parcerisa, L. (2017). A difficult relationship: Accountability policies and teachers. In M. Akiba & G. K. LeTendre (Eds.), *International Handbook of Teacher Quality and Policy* (pp. 241-254). Routledge.

Vindrola-Padros, C., Ledger, J., Barbosa, E. C., & Fulop, N. J. (2022). The Implementation of Improvement Interventions for "Low Performing" and "High Performing" Organisations in Health, Education and Local Government. *International Journal of Health Policy and Management*, *11*(7), 874-882.

Walgrave, S., & Dejaeghere, Y. (2017). Surviving Information Overload: How Elite Politicians Select Information. *Governance*, *30*(2), 229-244.

Wilson, J. Q., & Rachal, P. (1977). Can the government regulate itself? *The Public Interest*, 3-14.

Woo, S., Lee, S., & Kim, K. (2015). Carrot and stick?: Impact of a low-stakes school accountability program on student achievement. *Economics Letters, 137*, 195-199.

Appendix

Table A1: Summary statistics

	(1) Obs	(2) Mean	(3) SD	(4) Min	(5) Max
Panel A: performance turnaround					
Investigated upheld complaints (#)	1064	5.347	5.904	0.000	44.000
not accepted for investigation (#)	1064	6.553	6.183	0.000	39.000
Panel B: tunnel vision					
awaiting assessment completion (#)	1041	140.547	188.934	0.000	1448.750
awaiting public funding (#)	1041	31.482	43.057	0.000	421.417
discharge dispute (0/1)	1064	0.851	0.357	0.000	1.000
ease of finding info & advice on support (%)	1028	51.193	8.673	29.800	95.300
satisfaction with care services (%)	1025	89.840	2.926	68.200	98.600
impact of social care on quality of life (%)	1027	91.327	3.007	70.400	97.500
encouragement and support as carer (%)	1048	29.421	5.191	13.000	46.100
satisfaction with social services support (%)	1048	66.042	10.150	35.700	89.500
Panel C: leadership attention					
key decisions (#)	530	1.232	1.765	0.000	12.000
non-key decisions (#)	530	1.311	3.112	0.000	24.000
Panel D: organizational centralization					
audit days (#)	642	99.004	97.467	0.000	665.000
Panel E: core competencies					
all workforces (#)	1029	254.289	219.232	0.000	1675.000
workforces in frontline (#)	1019	122.589	115.710	0.000	808.000
workforces in supervisor (#)	1019	40.763	39.489	0.000	343.000
workforces in professional (#)	1019	41.075	29.010	0.000	257.000
Panel F: managerial turnover					
turnover of the director of adult care (0/1)	1023	0.319	0.466	0.000	1.000
Panel G: population demographics					
rate of health problems in the population (%)	1048	0.129	0.024	0.000	0.364
adult rate in the population (%)	1048	0.802	0.050	0.276	0.993

Table A2: Data sources

Variable	Data source
Panel A: performance turnaround	
Investigated upheld complaints (#) not accepted for investigation (#)	1. The Ombudsman's case management data: https://www.lgo.org.uk/
Panel B: tunnel vision	
awaiting assessment completion (#) awaiting public funding (#) discharge dispute (0/1)	2. Delayed transfers of care: https://www.england.nhs.uk/statistics/statistical-work-areas/delayed-transfers-of-care/
ease of finding info & advice on support (%) satisfaction with care services (%) impact of social care on quality of life (%)	3. Adult social care survey: https://digital.nhs.uk/data-and-information/publications/statistical/personal-social-services-adult-social-care-survey
encouragement and support as carer (%) satisfaction with social services support (%)	4. Personal social services survey of adult carers: https://digital.nhs.uk/data-and-information/publications/statistical/personal-social-services-survey-of-adult-carers
Panel C: leadership attention	
key decisions (#) non-key decisions (#)	5. The decisions page of each council: councils' websites
Panel D: organizational centralization	
audit days (#)	6. The internal audit plans page of each council: councils' websites
Panel E: core competencies	
all workforces (#) workforces in frontline (#) (e.g., social workers responsible for assessment and care plan) workforces in supervisor (#) (e.g., team leaders) workforces in professional (#) (e.g., occupational therapists)	7. Personal social services: staff of social services departments: https://digital.nhs.uk/data-and-information/publications/statistical/personal-social-services-staff-of-social-services-departments
Panel F: managerial turnover	
turnover of the director of adult care (0/1)	8. Municipal yearbook: British library
Panel G: population demographics	
rate of health problems in the population (%) adult rate in the population (%)	9. Adult population survey: https://www.nomisweb.co.uk/sources/aps

Multiple treatments

As noted in the article, several councils receive two or even three public interest reports related to adult social care during our study period. Six of these "multiple treatments" are for differing kinds of failure (assessment and care planning, charging and direct payments, and safeguarding), while seven cases involved second notifications for repeating the same type of failure.[20] To account for these two types of multiple treatments in our DiD estimation, we implemented two specific strategies.

First, for councils censured in different subcategories, we refined our analysis to the council–service category–year level. The treatment variable equals 1 for a given council–service category in all years following the first censure in that category, and remains 0 otherwise. Other service categories within the same council remain untreated during this period. Table A3 reports the estimates: Columns 1 and 2 use the number of upheld complaints as the outcome variable, controlling for local authority–category fixed effects and year fixed effects. The results remain consistent with the main council–year level analysis reported in Table 2 in the main article. Notably, the council–service category–year level analysis allows us to disaggregate the estimated effects by service category, providing insight into how each category

[20] As noted in the main text, the original count of multiple-treated councils is 16. However, three of these councils were excluded during the CEM matching process. Therefore, only 13 multiple-treated councils are included in the analysis.

responds to the issuance of a public interest report in terms of maladministration. Figure A3 (later in this appendix) shows that, while all three categories exhibit clear reductions in upheld complaints following a censure, the effect is most pronounced in the safeguarding category, whereas the effect in assessment and care planning is relatively weak.

Table A3: Robustness check: Effects of low-stakes accountability on performance turnaround at council-category level analysis

	(1)	(2)	(3)	(4)
	Investigated upheld cases		Not accepted for investigation	
Panel A				
Post report releasing	-0.718**	-0.741***	0.239	0.228
	(0.279)	(0.280)	(0.235)	(0.225)
Panel B				
Year 0	-0.529**	-0.507**	0.256	0.263
	(0.252)	(0.257)	(0.171)	(0.166)
Year 1	-0.778***	-0.816***	0.166	0.152
	(0.281)	(0.272)	(0.223)	(0.221)
Year 2	-0.750**	-0.765**	0.340	0.343
	(0.318)	(0.324)	(0.246)	(0.229)
Year 3	-0.816	-0.875*	0.192	0.155
	(0.524)	(0.517)	(0.385)	(0.383)
Local authority-category FEs	✓	✓	✓	✓
Year FEs	✓	✓	✓	✓
Demographic control		✓		✓
Empirical method	CEM+DiD	CEM+DiD	CEM+DiD	CEM+DiD
Observations	3,192	3,144	3,192	3,144

Notes: The unit of observation is the local authority-category × year. Post report releasing is a dummy that is 1 if local authority-category has received the report, and 0 otherwise. Standard errors clustered at the local authority level. *** p<0.01, ** p<0.05, * p<0.1.

Second, to address councils receiving more than one censure within the same service category, we extended the council–service category–year level analysis by creating a new variable, second report releasing, which

equals 1 in all years after the second censure in the same category and 0 otherwise. Given the limitations of Callaway and Sant'Anna's (2021) group-time average treatment effect estimator in handling multiple treatments, we instead adopted the standard stacked DiD approach commonly used in recent staggered DiD studies.[21] Table A4 presents the results: Column 1 includes both the first report releasing and second report releasing indicators, showing that the effect of the second censure is statistically insignificant. To further validate this finding, Column 2 interacts the two indicators to test for a triple-difference effect, which likewise proves insignificant. These findings suggest that, conditional on an initial censure, a subsequent censure in the same service category does not generate additional measurable effects.

Table A4: Robustness check: Effects of low-stakes accountability on performance turnaround at council-category level, considering for second censure

	(1)	(2)
	Investigated upheld cases	
First report releasing	-0.593***	-0.593***
	(0.182)	(0.182)
Second report releasing	-0.233	
	(0.222)	
First report releasing × Second report releasing		-0.233
		(0.222)
Report releasing wave		
× Local authority-category FEs	✓	✓
× Year FEs	✓	✓
Demographic control	✓	✓
Empirical method	✓	✓
Observations	22,296	22,296
R-squared	0.563	0.563

[21] Technical details of the stacked DiD approach can be found in Elston et al. (2025).

Notes: The unit of observation is the report releasing wave × local authority-category × year. First report releasing and second report releasing are dummies that are 1 if local authority has received the report, and 0 otherwise. Standard errors clustered at the local authority level. *** p<0.01, ** p<0.05, * p<0.1.

Table A5: Robustness check: Effects of low-stakes accountability on performance turnaround, never treated groups excluded

	(1)	(2)	(3)	(4)
	Investigated upheld cases		Not accepted for investigation	
Panel A				
Post report releasing	-1.093***	-1.117**	0.294	0.451
	(0.396)	(0.476)	(0.329)	(0.333)
Panel B				
Year 0	-0.832**	-0.787*	0.328	0.498*
	(0.358)	(0.419)	(0.268)	(0.264)
Year 1	-1.027***	-0.988**	-0.026	0.403
	(0.365)	(0.412)	(0.403)	(0.328)
Year 2	-1.503***	-1.569	0.548	0.469
	(0.579)	(0.980)	(0.359)	(0.348)
Year 3	-1.008	-1.123	0.325	0.434
	(0.667)	(0.744)	(0.504)	(0.578)
Local authority FEs	✓	✓	✓	✓
Year FEs	✓	✓	✓	✓
Demographic control		✓		✓
Empirical method	CEM+DiD	CEM+DiD	CEM+DiD	CEM+DiD
Observations	322	322	322	322

Notes: The unit of observation is the local authority × year. Post report releasing is a dummy that is 1 if local authority has received the report, and 0 otherwise. Post is a dummy that is 1 if the year is after the report releasing. Standard errors clustered at the local authority level. *** p<0.01, ** p<0.05, * p<0.1.

Table A6: Effects of low-stakes accountability on leadership attention and centralization

	(1) key decisions	(2) non-key decisions
Panel A		
Post report releasing	1.314*	-0.449
	(0.711)	(0.550)
Panel B		
Year 0	0.717	-0.406
	(0.710)	(0.576)
Year 1	1.550*	-0.446
	(0.797)	(0.553)
Year 2	2.125**	-0.157
	(0.925)	(0.669)
Year 3	0.862	-0.787
	(1.165)	(0.932)
Local authority FEs	✓	✓
Year FEs	✓	✓
Demographic control	✓	✓
Empirical method	CEM+DiD	CEM+DiD
Observations	530	530

Notes: The unit of observation is the local authority × year. Post report releasing is a dummy that is 1 if local authority has received the report, and 0 otherwise. Post is a dummy that is 1 if the year is after the report releasing. Standard errors clustered at the local authority level. *** p<0.01, ** p<0.05, * p<0.1.

Table A7: Effects of low-stakes accountability on centralization and managerial turnover

	(1) Audit days
Panel A	
Post report releasing	1.108
	(0.700)
Panel B	
Year 0	0.606
	(0.468)
Year 1	0.999
	(0.984)
Year 2	1.694*
	(0.918)
Year 3	1.135
	(1.462)
Local authority FEs	✓
Year FEs	✓
Demographic control	✓
Empirical method	CEM+DiD
Observations	951

Notes: The unit of observation is the local authority × year. Post report releasing is a dummy that is 1 if local authority has received the report, and 0 otherwise. Post is a dummy that is 1 if the year is after the report releasing. Standard errors clustered at the local authority level. *** p<0.01, ** p<0.05, * p<0.1.

Table A8: Effects of low-stakes accountability on turnover of the director

	(1) Turnover of the director
Panel A	
Post report releasing	0.034
	(0.093)
Panel B	
Year 0	0.099
	(0.101)
Year 1	0.030
	(0.117)
Year 2	-0.057
	(0.126)
Year 3	0.062
	(0.134)
Local authority Fes	✓
Year Fes	✓
Demographic control	✓
Empirical method	CEM+DiD
Observations	1,012

Notes: The unit of observation is the local authority × year. Post report releasing is a dummy that is 1 if local authority has received the report, and 0 otherwise. Post is a dummy that is 1 if the year is after the report releasing. Standard errors clustered at the local authority level. *** p<0.01, ** p<0.05, * p<0.1.

Table A9: Effects of low-stakes accountability on urgency

	(1)	(2)
	Investigated upheld cases	
	Δ Frontline workforces (year 0 vs. Prior 3-year) > National avg.	Δ Frontline workforces (year 0 vs. Prior 3-year) < National avg.
Panel A		
Post report releasing	-1.515***	-0.296
	(0.412)	(0.330)
Panel B		
Year 0	-0.958***	-0.325
	(0.368)	(0.400)
Year 1	-1.577***	-0.549
	(0.442)	(0.393)
Year 2	-1.566***	-0.241
	(0.538)	(0.350)
Year 3	-1.959**	-0.070
	(0.895)	(0.450)
Local authority Fes	✓	✓
Year Fes	✓	✓
Demographic control	✓	✓
Empirical method	CEM+DiD	CEM+DiD
Observations	951	963

Notes: The unit of observation is the local authority × year. Post report releasing is a dummy that is 1 if local authority has received the report, and 0 otherwise. Post is a dummy that is 1 if the year is after the report releasing. Standard errors clustered at the local authority level. *** p<0.01, ** p<0.05, * p<0.1.

Figure A1: Correlations of council characteristics between 2012 and 2013-2019

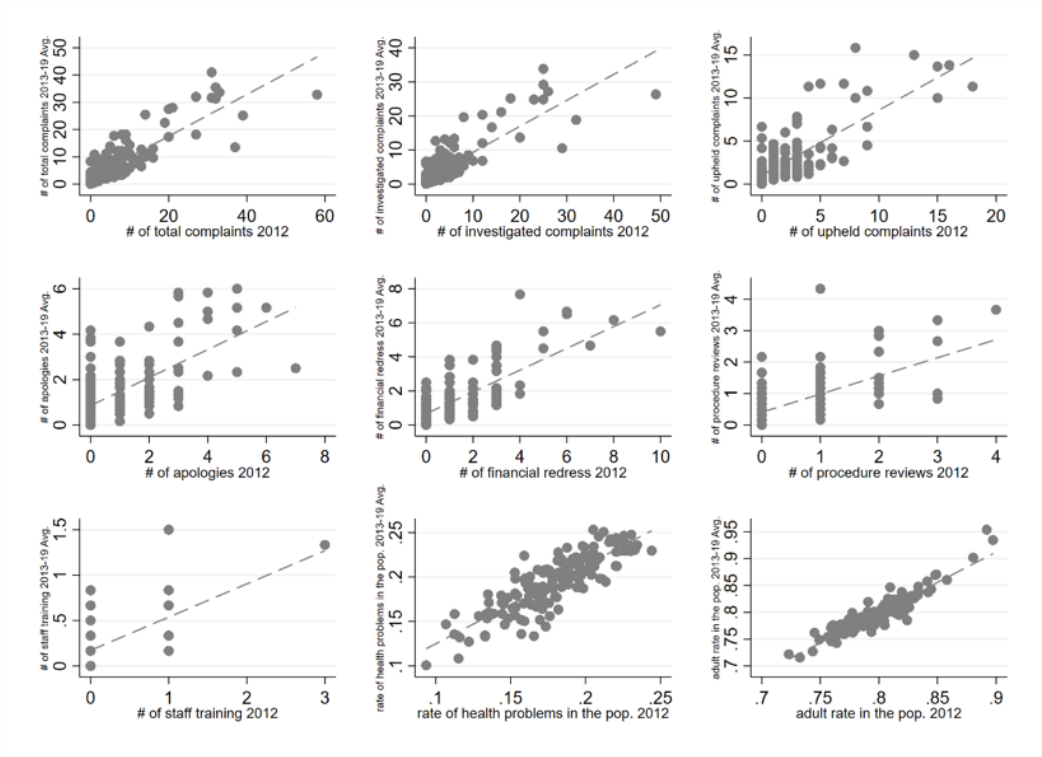


Figure A2: Event study - Performance turnaround in censured vs. non-censured councils at council-category level analysis

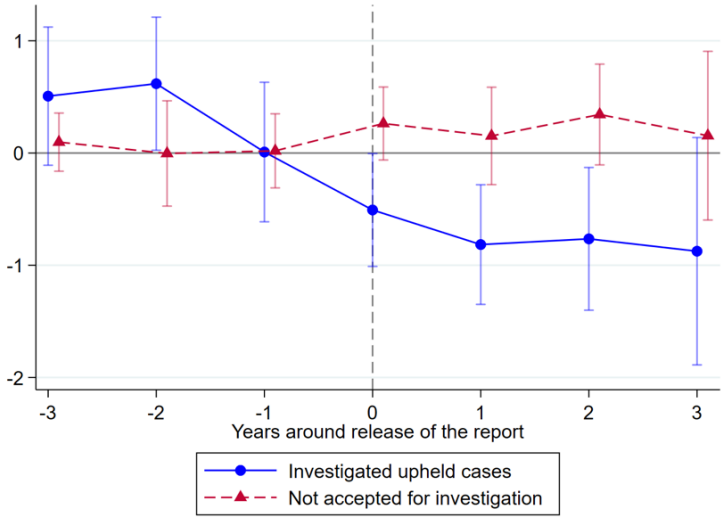


Figure A3: Event study - Performance turnaround in censured vs. non-censured councils across service categories

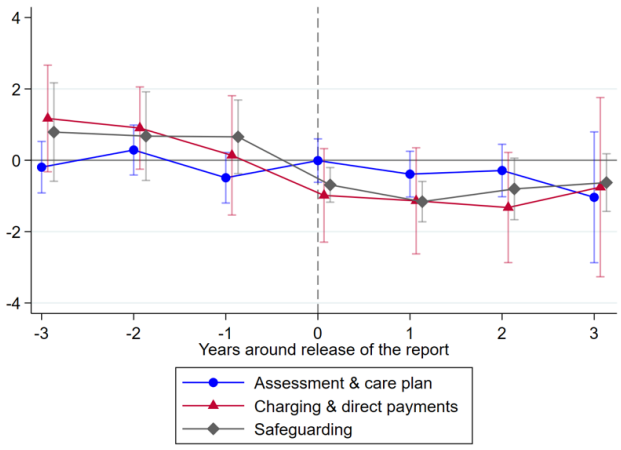


Figure A4: Event study - Performance turnaround in censured vs. non-censured councils, never treated groups excluded

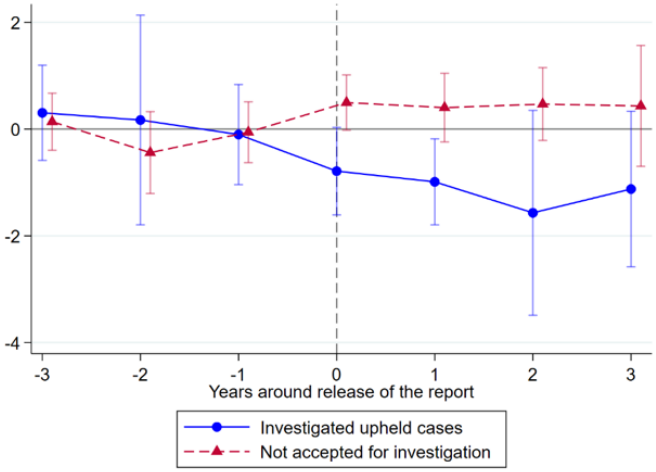


Figure A5: Event study - Core competencies in censured vs. non-censured councils across service categories