# Qualified Immunity and Police Employment

Paul Heaton
Santoro Giuggio

University of Pennsylvania Carey Law School

March 2026

Abstract: Qualified immunity (QI) protects state and local police officers from federal tort liability resulting from on-the-job misconduct that violates the rights of others. Defenders of QI have long argued that the doctrine plays an important role in maintaining police employment because without these protections officers would have difficulty performing their duties, and the U.S. Supreme Court essentially accepted this argument when in established the current framework governing QI in *Harlow v. Fitzgerald*. Yet there is little rigorous empirical evidence grounding this claim. In this paper we assess the impact of QI on police employment, exploiting the fact that Colorado and New Mexico recently essentially eliminated QI by legislating state law tort causes of action for constitutional violations for which QI does not apply, while the Nevada Supreme Court eliminated qualified immunity through a court decision. Using three different data sources, we conduct differences-in-differences and triple-differences analyses that contrast police employment levels over time, across states, and across professions. In a multiverse analysis, we show that conclusions regarding the impacts of QI are sensitive to modeling decisions, with some plausible models delivering evidence of large adverse impacts of eliminating QI and others showing positive impacts of QI reform. Combining information from hundreds of different empirical specifications, we demonstrate that there is no statistically discernable relationship between eliminating QI and police employment, and our estimates for all three states are sufficiently precise to statistically reject employment impacts of more than -9%. The repeal of QI in Colorado, New Mexico, and Nevada apparently did not produce a rush to the exits in police departments, calling into question one of the key assumptions underlying the current doctrine.

---

Guiggio: sgiuggio@penncarеylaw.upenn.edu; Heaton: pheaton@law.upenn.edu. We express appreciation to Misha Dworsky, Emily Leslie, Ross Miller, and seminar participants at RAND and the University of Pennsylvania for helpful comments. There are no external funding sources or potential conflicts of interest to disclose.

## I. Introduction

In recent years, numerous viral incidents involving law enforcement shootings of unarmed and non-threatening citizens have drawn public attention to the issue of police accountability. Amid numerous proposed policy reforms focused on accountability[1], one legal dimension that has received increased attention is qualified immunity (QI), a legal doctrine that shields officers from paying monetary damages for harms that violate the U.S. Constitution, including unjustified killings. Qualified immunity has been critiqued by progressives, originalists, and libertarians, and there have been notable recent efforts to get the Supreme Court to reconsider the doctrine.

Although much scholarly work explores the theoretical and historical underpinnings of qualified immunity doctrine, outside of the work of Joanna Schwartz, there has been relatively limited empirical work on the doctrine's effects. Proponents of qualified immunity, including the Supreme Court, argue that without qualified immunity, individuals would be unwilling to enter government service out of fear of being constantly sued for their official acts. Although there are reasons to doubt this claim—including from some of Schwartz's empirical studies—this proposition has yet to be rigorously tested empirically. In this paper, we exploit recent policy reforms in three states—Colorado, New Mexico, and Nevada—that eliminated qualified immunity as a defense in damages suits against officers for constitutional violations. Using a differences-in-differences research design, we estimate the impacts of qualified immunity reform on police staffing levels.

[1] As illustrative examples of recent research in this area, see Keyes and Keyes (2023) (civilian oversight), Kim (2025) (police body cameras), Sawyer and Widra (2026)) (data collection and transparency), and Zare et al. (2025) (training). For more critical perspectives of accountability reform measures, see, e.g. Rocha Beardall (2022) and Karakatsanis (2025).

The typical approach for a DD analysis is to choose a preferred model specification and then report robustness checks that vary design decisions around that baseline in the hope that they deliver similar results. However, for this question, there are several design decisions for which there is no clear answer as to what the "preferred" specification ought to be. For example, should we limit the analysis to years close to the present--which presumably were most similar in terms of external forces shaping crime rates than periods in the more distant past—at the cost of reducing our ability to account for trends in the data? Is it better to focus on states near to those that enacted reform, which arguably have a more similar geography and culture, or should the sample include all states? Many of these decisions involve trading off the benefits of a larger sample, which should offer greater statistical power to detect effects, with a sample where there is arguably better comparability between the treated and comparison group.

Rather than claiming one particular empirical specification is best, we instead conduct a multiverse analysis (Steegen et al. 2016) that evaluates a broad range of alternatives, exploring hundreds of different modeling design combinations across three separate datasets. Although at first glance it might seem doubtful that we could make meaningful claims about the effects of qualified immunity reform while remaining relatively agnostic about the statistical model, we demonstrate that this is not the case. We show that modeling design decisions do influence the final effects estimates and their standard errors, but for the most part not in easily predictable ways. When we incorporate data from a broad range of models and multiple datasets, we can draw a policy-relevant conclusion--we statistically reject decreases in police staffing of more than 9% in all three reform states. Whatever adverse impact qualified immunity reform has, it is not led to large-scale departures of police, as has been claimed by some reform critics. These results also challenge the premise that qualified immunity protections are essential to ensure an

adequate supply of government workers, one of the key assumptions used by the Supreme Court to justify its current approach to the doctrine.

Section II of the paper offers a primer on qualified immunity and describes the reforms implemented in Colorado, New Mexico, and Nevada. Section III describes our three primary data sources, and Section IV explains the model specification problem and our proposed solutions. Section IV presents the results of our analysis, including an exploration of which design decisions are most influential in shaping the ultimate conclusions produced from the differences-in-differences analysis. Section V concludes.

## II. Background

### a. Legal Origins

When an individual seeks compensation for harms caused by police officers, they can sue for damages under 42 USC § 1983. This provision was enacted as part of the Civil Rights Act of 1871 during the Reconstruction Era to allow individuals (particularly former slaves) to sue government officials for violations of their federal Constitutional rights in federal court. Despite this intent, judicial interpretation limited the law's effectiveness as a civil rights enforcement tool for many decades after its passage. It was not until the Supreme Court's ruling in *Monroe v. Pape*[2] that § 1983 was interpreted in a way that would allow it to become a major tool for victims of police-caused harms seeking to litigate police misconduct (Schwartz 2023).

Not long after *Monroe* opened the door to civil suits against police in federal courts, police liability was curtailed by the creation of the doctrine of qualified immunity. The doctrine first emerged in *Pierson v. Ray*[3] as a defense against liability for police whose improper conduct was

[2] 365 U.S. 167 (1961)
[3] 386 U.S. 547 (1967)

carried out in good faith. Years later, in *Harlow v. Fitzgerald*[4], the Court removed concern over officers' intentions from the qualified immunity analysis and replaced the good faith standard with the current "clearly established" standard for granting qualified immunity (Schwartz 2023).

Under current law, to prevail in a § 1983 case, a plaintiff must show that (1) their constitutional rights were violated and (2) that there was clearly established precedent that would put an officer on notice that their actions violated the plaintiff's rights. Courts have some leeway in deciding whether to deal with these prongs of the § 1983 analysis in order, or whether they can skip to the second prong and dismiss the case for lack of clearly established precedent right away.[5] It is this second prong that grants qualified immunity to an officer accused of misconduct so long as there is no prior case condemning the same conduct committed under similar circumstances. As the courts have given content to this standard and outlined just how similar a case of misconduct needs to be to a prior case to count as clearly established, the standard has become increasingly difficult (some would say nearly impossible) to meet (Schwartz 2023).

Take, for example, *Corbitt v. Vickers*[6]. In this case, police followed a suspect into the yard of an unrelated party, Amy Corbitt. Several children playing in the yard were ordered to get on the ground and held at gunpoint while the officers sought to take the suspect into custody. The children all complied, but when Corbitt's dog came into the yard, without any provocation, an officer fired at the animal several times and accidentally hit Corbitt's 10-year-old child who was lying on the ground about three feet away from the officer. The child suffered major injuries due to the gunshot wound but when Corbitt sued the police, qualified immunity protected the officers from liability. Despite a dissenting judge's argument that the officer's "conduct—discharging a

[4] 457 U.S. 800 (1982)
[5] *Baxter v. Bracey*, 751 F. App'x 869 (6th Cir. 2018), *Jessop v. City of Fresno*, 936 F.3d 937 (9th Cir. 2019).
[6] 929 F.3d 1304, 1325 (11th Cir. 2019)

lethal weapon at a nonthreatening pet that was surrounded by children—is plainly unreasonable*",* the appellate court determined that there was no case to clearly establish that an officer firing at a dog and accidentally hitting an innocent detainee constituted a violation of the Fourth Amendment.

Consider as well *Baxter v. Bracey*[7], where officers were granted immunity when they released their police dog on a suspect who was surrendering. The court found that loosing police dogs on a suspect sitting down with their hands up was not a clearly established violation, even though earlier case law noted that loosening dogs on a suspect who was lying down was improper. This case reflects how, under current law, minutiae like whether the victim was sitting or lying down can render substantially similar misconduct immune from suit[8]

**b. Debates: Scholarly and Popular**

The doctrine of qualified immunity is deeply contested in the public sphere and in academic circles. Disagreements cluster around both its legal basis and real-world effects as a public policy choice.

On the legal side of the debate, there are critics who dislike the doctrine's judge-made character and call into question its legal soundness on those grounds. Baude (2018), for example, argues that §1983 contains no textual basis for providing government officials with an expansive defense to liability like qualified immunity and that the doctrine cannot be derived from other legal principles either. He concludes that "the restrictions imposed by qualified immunity have been wrongly imposed by the Court, not implied by the statute or the common law. The Court's

[7] 751 F. App'x 869 (6th Cir. 2018)

[8] There are some cases in which the Supreme Court finds the underlying misconduct so egregious that there was no need for precedent to clearly establish that the police were acting improperly. However, these cases are very rare and tend to involve truly heinous conduct. See, e.g,.*Hope v. Pelzer* 536 U.S. 730 (2002) (finding that no prior case law was needed to demonstrate that it was illegal to leave a prisoner tied up outside in the southern sun for a whole summers day) or *Taylor v. Riojas* 592 U.S. 7 (2020) (denying the need for prior case law to establish that holding prisoner for a week in a cell covered in excrement violated the law).

crusade to enforce the doctrine of qualified immunity does not serve congressional intent or the rule of law." (Baude 2018). Other legal scholars also join Baude in challenging the lawfulness of the doctrine (Reinert 2023; Schwartz 2017b).

Prominent among those who defend the doctrine on legal grounds is the Supreme Court itself. Baude's article is a response to the legal rationales employed by the Court to justify its qualified immunity holdings. These include arguments that the doctrine has its roots in common law tradition that can be read into § 1983 and claims that the law of lenity and fair notice provide the legal basis for the doctrine (Baude 2018). Scholars as well have taken up the cause of defending the doctrine on a variety of legal grounds (Nielson and Walker 2017; 2020; Rosenthal 2020).

In addition to these legal debates, the policy implications of qualified immunity are heavily contested. Critics argue that the doctrine prevents victims from recovering compensation for violations of their rights ("Frequently Asked Questions About Ending Qualified Immunity," n.d.; "Qualified Immunity," n.d.; Schweikert 2020; Schwartz 2017a). The defense lets police off the hook for harms inflicted on victims, not because there was no harm, or no violation of the law or even because the officials in question acted in good faith. It lets officers off the hook for conduct that simply lacks precise legal precedent establishing that it is illegal. As the law currently stands, a court can find officers' actions illegal and grant them immunity anyway. In such cases, victims' harms will go unrectified and their rights unvindicated.

Critics also raise concerns about the incentives such broad immunity creates. If police officers or their departments cannot be effectively held liable for officers' misconduct, courts remove an incentive for law enforcement to internalize the harms they impose on others. In doing so, they mitigate the deterrent effect of litigation and encourage further police misconduct

(Carlos 2018; Crocker 2021). The issue of whether constitutional tort litigation actually has much of a deterrent effect on government behavior is disputed, with some scholars defending the efficacy of litigation (Gilles 2000; Karlan 2006), some disputing it (Armacost 2003; Levinson 2000), and others problematizing these accounts (Schwartz 2009). Despite this scholarly disagreement, the potential deterrent value of civil rights lawsuits remains a major reason for questioning qualified immunity as a policy choice.

Against these criticisms, defenders of qualified immunity raise the concern that without these protections the justice system will experience a number of dysfunctions. Without some form of qualified immunity, officials could be unjustly held liable for conduct when they lack fair notice that their actions are improper (Chapman 2023). An expansive reading of § 1983 without qualified immunity could also flood the federal courts with spurious civil rights cases and bankrupt police departments and their officers (Schwartz 2023).[9] While there are other scholarly works that discuss and dispute these concerns (Merrell 2021; Schwartz 2019), this paper is focused on a different (albeit related) fear of the consequences of weakening qualified immunity: the idea that the doctrine is necessary to maintain sufficient levels of police employment. Defenders of the doctrine worry that the threat of tort liability will deter potential officers from taking an already difficult public service job. This concern was invoked by the Supreme Court as a reason for creating the modern form of the doctrine in *Harlow v. Fitzgerald*[10]. It has also been proffered as a reason not to reform qualified immunity in a variety of scholarly and popular contexts, sometimes on a limited evidentiary basis.

In a significant and influential body of scholarly work in this area, Joanna Schwartz (2023) argues that separate and apart from qualified immunity, restrictions on awarding attorneys' fees

[10] 457 U.S. 800 (1982)

to successful plaintiffs, the heightened pleading standard under cases like *Bell Atlantic v. Twombly*[11] and *Ashcroft v. Iqbal*[12], and the actual substance of judicial interpretations of the Fourth Amendment all help to shield police from being liable for their unconstitutional actions. She investigates the frequency with which officers actually face financial consequences as a result of litigation against them and demonstrates that police officers are almost always indemnified by their employers and therefore do not bear the costs of litigation individually (Schwartz 2014). This finding calls into question whether police officers, acting as rational agents, would actually change their professional path in response to changes in qualified immunity doctrine.

There is literature that cuts against this implication of Schwartz's work. A few studies use survey-based methods to show that police officers and potential recruits are concerned with future legal liability when making career decisions (Rossler and Scheer 2024; Santos et al. 2024). And some legal commentators have argued that state-level qualified immunity reforms have already had an adverse impact on police employment (Averett 2023). Nonetheless, there is a lack of clear and direct empirical evidence to help policymakers understand whether qualified immunity has a significant impact on police employment and recruiting. Using data derived from recent reforms in Colorado, Nevada, and New Mexico, this paper provides the first such evidence.

Providing this kind of evidence is all the more urgent given the variety of speculative claims that abound in popular discourse about the doctrine's effects on police employment. Senator Tom Cotton (2021), in the wake of the killing of George Floyd, released a statement about the importance of qualified immunity, which claimed that "[i]n all likelihood, the practical result of

[11] 550 U.S. 544 (2007)
[12] 556 U.S. 662 (2009)

eliminating or significantly curtailing qualified immunity would be fewer police." Similar statements can be found in various news outlets discussing possible reforms to qualified immunity at the federal or state levels (Rosen 2020; Balsamini 2021; Drysdale 2025; Fisher 2025; Keene 2023; McDonald 2021; Sherry 2022). For example, an official from the Fraternal Order of Police claimed in an Op-Ed that Colorado's qualified immunity reform was responsible for officers "retiring early, leaving the profession altogether or moving to other states to continue their careers" (McDonald 2021). The extent to which these effects actually materialize is clearly a matter of significant public concern.

**c. The Critics Win Out: QI Repeal in Colorado and New Mexico**

Following the death of George Floyd, many states began considering police reforms to prevent future lethal encounters between citizens and police (Gildin 2024). In a significant change to the legal landscape of qualified immunity doctrine, Colorado and New Mexico both passed police reform bills that repealed qualified immunity for state and local police. These laws did so by creating a state law cause of action that allows victims of police injury to sue under the state constitution for the same kinds of violations of their rights that they would otherwise have to turn to § 1983 to vindicate. By side-stepping the federal statute, these states create a means for victims to have legal recourse against the police for which the federal doctrine of qualified immunity is statutorily inapplicable.

The law repealing qualified immunity in Colorado, the Law Enforcement Integrity Act, was passed in June 2020 and became effective upon signature by the governor. The law only applies to state and local police, creating a state law cause of action that could be used to sue law enforcement officers for violations of the state constitution without qualified immunity as a

defense. The law requires police departments to indemnify officers, but allows a limited amount of up to $25,000 of the judgement against them to be borne by the officers who failed to act reasonably or in good faith. The bill includes other police reform measures as well, such as requiring body cameras for law enforcement officers, banning the use of certain types of force, and various reporting and data collection requirements (Enhance Law Enforcement Integrity Act 2020).

Effective July 1, 2021, the New Mexico Civil Rights Act created a cause of action under the New Mexico state constitution that could be used by victims of police misconduct to recover damages. The law empowers victims to sue public bodies at the state and local level for violations of their rights and expressly disallows qualified immunity as a defense. It set a $2 million cap on damages awarded in misconduct suits and allowed attorneys' fees to be awarded to victims (New Mexico Civil Rights Act 2021).

**d. Judicial Reforms at the State Level**

In addition to the legislative repeals of qualified immunity in Colorado and New Mexico, the state supreme courts of Montana and Nevada have created a state law workaround to federal qualified immunity in their jurisdictions. In 2002, the Montana Supreme Court held that litigants could sue for damages to gain compensation for violations of their rights by government officials under the state constitution. They also found that qualified immunity was not an available defense to such claims. The case, *Dorwart v. Caraway*[13], which involved unreasonable search and seizure claims against law enforcement, is still good law in Montana. Similarly, in the 2022

[13] 58 P.3d 128 (2002)

case *Mack v. Williams*[14], the Nevada Supreme Court held that suits for damages under the state constitution were not entitled to qualified immunity protections.

Because of these cases, law enforcement in these states can be subject to liability for violating the state analogues of the Fourth Amendment right to be free from unreasonable search and seizure without benefiting from §1983's qualified immunity protections. Although the Montana reform predates our sample period, in light of this judicial reform we evaluate Nevada as a treatment state alongside Colorado and New Mexico.

## III. Data

For our analysis, we make use of three main data sources: the American Community Survey (ACS), the Annual Survey of Public Employment and Payroll (ASPEP), and the Law Enforcement Employment Series (LEE)[15]. Each data source has particular strengths and weaknesses which allow us to run complementary analyses.

The ACS is a large-scale demographic survey of 1% of U.S. households conducted annually by the US Census Bureau. Employed respondents indicate their current profession in the survey, so the ACS can be used to obtain annual counts of workers by profession in each state. Because the ACS surveys the general population, it allows us to compare employment in policing with employment in other professions (e.g. security), including non-government employment. The survey also measures several useful outcomes beyond just headcounts of officers, like their wages and hours worked. A key disadvantage is the nature of the sampling scheme—with a 1% sample, even for a profession with 10,000 employees in a state in a given year, on average we'd

[14] 522 P.3d 434 (2022)

[15] We also explored using the Quarterly Census of Employment and Wages (QCEW), but ultimately found this unsuitable for our purposes because of variable and in some cases high non-reporting rates for state government employees due to their exclusion from the regular unemployment insurance system in many states.

observe only 100 ACS respondents, so we'd expect the ACS to produce a fair bit of statistical noise in measuring headcounts for specific professions.

The LEE is a part of the FBI's Uniform Crime Reporting program. It also provides annual counts of law enforcement employees by agency, including both civilian employees of law enforcement agencies and sworn officers. Because agencies are asked to report counts of employees, it should operate like a census, eliminating the sampling error problem that comes from the ACS. Moreover, as an agency-level survey, it has more granular geographical coverage than other data sources. Its main disadvantages are twofold. First, because it only contains information about police employment but not other professions, it cannot be used for triple-difference analyses. Second, the employment data are not meaningfully audited and there are no real consequences to reporting agencies from providing inaccurate staffing information, so these data may be less accurate than others data sources.[16] Many agencies, particularly smaller ones, don't report staffing information to the FBI at all or report only in a limited set of years, and in those that do, it is not uncommon to observe staffing levels that fluctuate implausibly–for example, showing increases of 50% or more in staffing levels in a single year in a larger agency. Below we discuss various approaches to deal with these LEE data quality limitations.

Each March, the Census Bureau administers the Annual Survey of Public Employment and Payroll (ASPEP) to state and local governmental entities to get information about headcounts across various job roles. Like the LEE, the ASPEP includes information about all personnel, alleviating the sampling problem. Unlike the LEE, ASPEP data are subjected to several data quality checks by Census personnel, so they raise fewer data quality concerns. The ASPEP data also include series on other categories of government (but not private) employment, furnishing

[16] Chalfin and McCrary (2018) discuss the problem of measurement error in the UCR data in more detail.

an opportunity for triple-differences analyses. Although some geographic disaggregation is possible using the ASPEP, for this paper, we focus on aggregate headcounts across all state and local agencies.[17]

## IV. Analytic Approach

If eliminating QI impacts police recruiting and retention, we would expect law enforcement agencies in Colorado and New Mexico to report fewer personnel after these states' law changes than would have been the case absent a law change. We measure such changes in a differences-in-differences (DD) framework. To illustrate the DD approach, Figure 1 plots the aggregate annual count of officers in Colorado from the ASPEP data, along with two potential comparison series—police officers in Arizona and firefighters in Colorado. One sensible possibility would be to use police staffing levels in other jurisdictions to project the counterfactual for what would have happened in Colorado absent a law change. This approach allows us to adjust for general trends that affect the policing profession as a whole over this time period. Given that police staffing levels rose in Colorado post-reform while they fell in Arizona, the comparison state, a simple cross-state DD using the data from Figure 1 would suggest that QI reform did not adversely impact police staffing levels.

Alternatively, we might look within Colorado and use other professions to construct the counterfactual. This form of DD estimate has the advantage of not relying on out-of-state comparisons; however it does not permit us to account for potential unobserved policing-specific shocks in particular years. As shown in Figure 1, if we use Colorado firefighters as a comparison

[17] The U.S. Census Bureau (2025) cautions against building time-series for individual localities using these data.

group, there is some indication that the QI reform may have harmed police recruiting and/or retention, as firefighter staffing levels seemingly rose more than police staffing levels.

A potentially more robust approach is to combine both comparisons in a triple-differences (DDD) framework. For a DDD analysis, the unit of observation is a state/year/profession combination, the treated group is policing in Colorado, New Mexico, or Nevada after their respective reforms, and the comparison groups include other professions, other states, and other years. The DDD analysis permits inclusion of a full set of profession by year, profession by state, and state by year fixed effects, which account for unobservable state-specific factors that affect employment (e.g. other general law changes) and well as factors such as the George Floyd protests that occur at particular points in time and disproportionately affect particular professions like policing.

Table 1 presents DD and DDD estimates, focusing on the ASPEP state-level panel covering 2005-2024. Here and in all the analysis that follow, we separately estimate impacts for Colorado, New Mexico, and Nevada; given the important differences in the specifics of how QI reform was implemented in each state, we are unwilling to assume a common effect across these states.

One issue surfaced in Table 1 is that we can draw differing conclusions about the effects of the law depending on whether we credit the DD or the DDD analysis. In the DD analysis, we find statistically significant increases in police staffing after immunity reform in Colorado and Nevada, but statistically significant decreases in New Mexico. In the DDD analysis, the impacts for Colorado reverse sign and become negative and significant while the estimates for New Mexico and Nevada are no longer statistically significant. However, here it seems unclear whether we should favor the DD or DDD approach—the DDD approach makes some additional assumptions about the data generating process, but in doing so, it yields estimates that are robust

to some confounding factors (e.g. other state-specific policy changes that occurred around the time of the QI reform) that would bias the simpler DD estimates.

More broadly, there are other design decisions that create indeterminacies regarding the baseline model. For example, what time window should we use for the analysis? If we use a narrower time window (e.g. 2015-2024), we avoid extrapolating from more temporally distant years where conditions may have been fairly different from those that prevailed around the time of QI reform. However, with a narrower time window we have less data and less ability to identify the amount of natural variation in the outcome that occurs over time absent policy change. A similar issue arises with respect to the geographic scope of the sample. Limiting the analysis to states close to Colorado and New Mexico (e.g. adjacent states) ensures that we focus attention on areas that are arguably most similar to the treated states in terms of physical geography, culture, or illicit market structure. On the other hand, including all states avoids the need to make potentially arbitrary or motivated decisions about how to narrow the sample, and enables us to have a broader set of control groups with which to identify effects. As another example, in both Colorado and New Mexico, the actual reforms went into effect mid-year. Because these are annual data, should we treat the year that the reform went into effect as a treated year, a control year, something in between, or completely separately? Each approach is defensible, and the choice may depend on typically unstated assumptions regarding treatment effect dynamics or the value of maximizing statistical power.

Most DD papers choose one baseline specification that the authors argue is the most appropriate one, and then explore robustness of the results across specification changes relative to that baseline. We adopt a different approach. Rather than trying to claim that one of the specifications is best, we estimate a range of different DD and DDD specifications that vary

these design decisions, and then consider what the combined results tell us, incorporating information across all available specifications.[18] Relative to the typical approach, this approach is likely to provide greater robustness to model uncertainty, and better address the problem of researcher degrees of freedom (Simmons et al. 2011) or forking paths (Gelman and Loken 2014). The downside of this approach is that it is potentially computationally intensive and may yield less definitive statements about the relationship of interest. However, if this approach does yield wider confidence intervals than the traditional approach, we argue that is in fact a better description of the true uncertainty around the effects of QI reform that incorporates not only statistical uncertainty but also model uncertainty.

Table 2 describes the variations in specifications, or design decisions, that we consider for each dataset. As noted above, we evaluate the use of a DD versus a DDD model, what year to start that dataset, which states to include in the control group, and how to define the timing of treatment. We also vary whether the data are population weighted—applying population weights arguably renders the estimates more broadly representative of the U.S. as a whole and might help with precision—and the comparison industries considered in the DDD analysis. For the ASPEP data, we consider specifications in both logs and levels[19], and we consider specifications that exclude a set of states the implemented other types of immunity reform short of full elimination of qualified immunity during the sample period.[20]

[18] Leamer (1983) argued for such an approach in a classic paper. Although not common in law and economics, the macroeconomic empirical growth literature has embraced this approach. See, for example, Levine and Renelt (1992), Sala-I-Martin (1997) and Sala-I-Martin et al. (2004).

[19] Because the ACS has itself become an important component for most annual state population estimates, denominating our ACS counts by state population to form a per capita measure could potentially introduce unknown sources of division bias for this dataset. For the LEE data, the relevant denominator would be the population served by each law enforcement agency, which isn't readily available for some LEAs.

[20] The excluded states are California, Connecticut, Maryland, Michigan, and Montana; Appendix XX documents the relevant reforms in each state.

For the LEE, we consider a number of design decisions that might potentially improve data quality. First, we consider a variant that drops very small agencies (fewer than 10 officers), which are less likely to report to the LEE and which have fairly noisy headcounts when measured in percentage terms. Our widest data window spans 2003-2024, a period of 22 years, but some agencies report data in substantially fewer than all 22 years. Our second data variant excludes, respectively, agencies that report fewer than half of the years or agencies that report fewer than 20 years. Because the analysis relies on state/year level variation, we also consider a variant that excludes states and years where more than half of the state's agencies do not report to the LEE. Finally, we consider a variant that excludes agencies with more than 10 officers that report changes in officer staffing levels of more than 50% in consecutive years; such large changes seem likely to reflect data entry errors rather than true dramatic increases in staffing levels.[21]

For each dataset, we consider all possible combinations of the listed design decisions, yielding 144 specifications for the ACS, 192 for the ASPEP, and 816 for the LEE. Under standard large-sample regression theory, each specification's coefficient estimate is approximately normally distributed. We therefore aggregate these specification-specific normal distributions into an equally weighted mixture of normals for each dataset. The resulting combined distribution characterizes our estimated effect size over a broad range of alternative identification assumptions and provides a basis for constructing omnibus confidence intervals.

[21] As an example of the types of errors that occasionally occur in the LEE data, the Provo (UT) Police Department reported between 100 and 109 officers in every year between 2012 and 2019. In 2020, it reported 198 officers. Then in 2021 the officer count was back to 98, and it remained between 100 and 115 for the rest of the sample.

## V. Results

### a. Basic Results

Figure 2 plots the coefficient estimates and 95% confidence intervals for the 144 different ACS specifications we considered. The 95% confidence intervals are based on standard errors clustered on state or state by industry. Important details of reform differed across the three states—for example, New Mexico implemented a damage cap where Colorado did not, Colorado allowed for individual plaintiffs while New Mexico did not, and Nevada implemented its reform through a judicial decision rather than legislation—and so in our view it is not appropriate to aggregate all three reform states together; instead we present separate estimates for each state.

Panel A shows the estimates for Colorado. The most notable pattern in Panel A is that it shows both negative and "statistically significant" and positive and "statistically significant"—at least if one considers only a single hypothesis test—coefficients, depending on specification. Similar patterns are apparent in Panels B and C for New Mexico and Nevada. For Colorado and New Mexico, the effects estimates appear roughly equally distributed around a null effect; for Nevada, the estimates skew a bit more negative overall, but there are still a substantial minority of estimates that are positive and "statistically significant"[22]. Consistent with Table 2, it is clear from Figure 2 that one can draw nearly any quantitative conclusion about the effects of qualified immunity reform using the ACS depending on the choice of sample and econometric specification. Stated differently, it would be possible for a motivated researcher to present plausible specifications that support the conclusion that QI reform harms police hiring, or, alternatively, that it actually counterintuitively improves police hiring.

[22] This qualitative conclusion would not change even if we adjusted for multiple comparisons using a conservative method such as the Bonferroni correction.

Figures 3 and 4 present analogous results for the ASPEP and LEE. For the ASPEP, there are many estimates with 95% confidence intervals that are fully negative for Colorado and a handful of positive ones; for New Mexico, the estimates are fully negative or statistically indistinguishable from zero, and for Nevada the estimates largely have 95% confidence intervals that include zero or are fully positive. The figure also reveals substantial differences across specifications in the precision of the estimates, with some specifications yielding confidence intervals that are 5x-10x narrower than others for the same state. For the LEE, which features much larger samples and more specifications but no triple-differenced estimates, we observe seemingly precisely estimated and for the most part statistically indistinguishable from zero estimates for Colorado. Confidence intervals are wider for New Mexico and Nevada, and for new Mexico there appear to be a substantial number of estimates with confidence intervals that are fully negative, while for Nevada we observe both fully positive and fully negative confidence intervals.

**b. Combined Estimates**

Standard regression assumptions imply that the coefficient estimate from each specification is approximately normally distributed for large samples. To summarize evidence across specifications, we combine these specification-specific normal approximations into an equally weighted mixture distribution[23]. This mixture represents the distribution of effect sizes across the full set of modeling choices.

[23] We have chosen equal weights so as to be agnostic about which specification is preferred. One could of course upweight particular specifications—e.g. giving greater weight to DDD over DD estimates--if one had prior beliefs about which specification choices are preferred. Indeed, the usual approach of choosing a main specification and doing robustness checks can be thought of as a version of this analysis which sets the weights for most specifications to zero and uses somewhat ad hoc methods for combining information across specifications.

Figure 5 plots the corresponding distributions for each state and dataset. Several patterns are notable. First, the ACS yields much more diffuse distributions than the other two datasets; which is perhaps not unexpected given that it is a sample rather than an ostensible census. Second, many of the distributions are multi-modal, which suggests that there are certain design choices that exert a strong influence on the realized estimates. Third, despite incorporating information from a much wider set of samples and identification assumptions than is typical in the law and economics literature, the distributions, at least for the ASPEP and LEE, appear sufficiently narrow to exclude practically meaningful effects, at least in some states. Table 3 reports the 95% confidence intervals associated with the distributions in Figure 5.

One potential concern with the aggregation approach above is that it doesn't explicitly account for any covariance between estimates that arises because the underlying specifications rely on the same data. As a check for whether such covariance matters, we implemented a bootstrap procedure where we randomly sampled states with replacement to generate a new dataset, reran the full set of specifications on the bootstrap sample, and repeated this process 1,000 times. This approach allows us to make minimal assumptions regarding the distributional characteristics of the estimates across specifications; instead, we can simply describe the full joint distribution of estimates. The bootstrap confidence intervals, also reported in Table 3, are very similar to those obtained using traditional analytic methods.

To more fully explore which design decision matter and by how much, we estimated dataset/state level regressions where the unit of observation is a model specification, the outcome is the coefficient estimate, and the primary explanatory variables are indicators for various design decisions—for example, whether the data are weighted or whether the sample begins at 2010 or 2015. Table 4 reports the results of these analyses. Panel A documents the relationship

between design decisions and the point estimates of effect sizes, while Panel B focuses on standard errors.

Several notable patterns are apparent in Panel A. First, design decisions matter—for the ACS, for example, weighting by population increases the estimates by 4-7 percentage points, depending on state, and limiting the sample period to 2015 shifts the estimates by 3-11 percentage points. Second, it is generally the case that the impacts of a particular design decision vary in sign across states. As an example, using the less expansive definition of treatment increases the ACS estimates for Nevada but decreases the estimates for Colorado and New Mexico; increases Colorado's ASPEP estimate while leaving the other two states unchanged; and increases the LEE estimates for Colorado and New Mexico while decreasing the estimate for Nevada. Third, across datasets, the impact of the same design decision is often variable. As an illustration, limiting the analysis to the Census Mountain division reduces all three states' coefficients for the LEE, increases all three states' coefficients for the ASPEP, and has a mixed impact for the ACS. The one exception to this last pattern is population weighting, which increases the estimates for nearly all states and datasets.

The results for precision (Table 4, Panel B) are more consistent across states and datasets. Some design decisions, such as how to define the start of treatment, whether to exclude states with other types of immunity policy reforms, and the various LEE data quality measures have little discernable effects on the precision of the estimates. Moving from a DDD to a DD specification improves precision, as does limiting the sample to more recent years and including all states rather than only the Census Mountain Division states. Population weighting has an ambiguous effect, improving precision for the ACS, having no impact for the ASPEP, and reducing precision for the LEE.

Taken as a whole, these patterns support several conclusions about the role of the model specification in shaping our conclusions about the impacts of qualified immunity reform. First, the modeling decisions matter—the effect sizes change by statistically significant and practically important amounts depending on particular modeling choices. Taking a wholistic view that considers a much wider range of modeling decisions than is typical helps to reveal modeling uncertainty as an important additional source of uncertainty beyond sampling uncertainty in assessing real-world policy effects. Second, however, with the possible exception of population weighting, there isn't an obvious way to "stack the deck" in favor of a particular conclusion—specific modeling choices often trade off "better" results for one state and dataset for "worse" results for others. Third, Table 4 suggests that the typical approach of changing only one or a few aspects of the econometric specification for robustness checks may be overly limiting, as it is possible to get more substantial deviations from the baseline estimates by simultaneously altering many design decisions. Because we explore all possible combinations of the various design decisions, we are able to more fully characterize the true degree of model uncertainty than is typical.

### c. Meta-Analysis

For a given dataset, the various specifications examined above all rely on the same underlying data, which is why it is not appropriate to regard each collection of estimates as a series of independent draws. However, this is not true when we conduct comparisons across datasets. Because each dataset seeks to measure the same underlying causal relationship, we could combine information across the three datasets to potentially obtain tighter confidence

intervals for the effect size. Combining separate estimates of the same underlying causal relationship is the domain of meta-analysis.

In Table 5, we report meta-analytic estimates that incorporate information across all three datasets. We consider two approaches for formulating the inputs into the meta-analysis. First, we find the best fit normal distribution to each distribution shown in Figure 3, and then use those distributions' means and standard errors in our meta-analytic calculations. Alternatively, we identify the normal distributions that would yield the same 95% confidence intervals as those given in Table 3 and use those for the meta-analysis. Both approaches yield similar conclusions.

Even though most of the individual confidence intervals reported in Table 3 include practically meaningful increases and decreases from qualified immunity reform, when the estimates are pooled, we are able to statistically reject fairly modest decreases in police staffing levels. For Colorado and Nevada, we can statistically reject decreases of more than 6% in police staffing; for New Mexico, we can reject decreases of more than 9%. Although the data are not inconsistent with a small adverse effect of QI reform on police staffing, the predictions of some opponents of QI reform, who argued that limiting immunity would create significant adverse impacts on police recruiting, are not borne out in the data.

### d. Other Outcomes

[Add analysis of earnings]

[Add analysis of officer characteristics--unsworn, female, non-White, etc.]

## VI. Conclusions

Qualified immunity doctrine rests on many largely untested assumptions. One significant one, embraced for many years by the U.S. Supreme Court, is that the protections of qualified

immunity are necessary to ensure a robust supply of workers for government positions that require discretionary exercise of state power. Using recent reforms in Colorado, New Mexico, and Nevada, and drawing on three distinct data sources, we find little support for that assumption, at least in the context of modern policing. Across a large multiverse of plausible differences-in-differences and triple-differences specifications, the estimated effects of qualified immunity reform on police employment vary in magnitude and sometimes in sign. But when considering results together, the overall pattern is clear: eliminating qualified immunity did not produce a substantial decline in police staffing in any of the three reform states.

Of course, our analysis doesn't conclusively resolve the normative debate regarding qualified immunity; there remain important contested questions regarding its historic roots, deterrent effects, and impacts on fairness and racial justice. Moreover, repeals of qualified immunity may have impacts on other margins not considered here such as officer quality, morale, and local government finances. At a minimum, however, we show that the most dire claims regarding the adverse impacts of eliminating qualified immunity have not been realized in those states that have embraced reform.

Beyond its contribution to an important substantive legal question, this paper also demonstrates the utility of the multiverse analysis, a method that has not yet gained widespread application in legal and economic scholarship. Rather than privileging a single preferred specification, we show that reasonable design choices can materially affect point estimates and levels of statistical significance. Here the conventional approach centered on one baseline model and a more limited set of robustness checks could easily create a misleading sense of certainty. The multiverse approach offers a more candid account of the uncertainty surrounding causal inference. The fact that we can adopt this broader perspective without sacrificing the ability to

draw policy-relevant conclusions suggests that the multiverse analysis may be worth exploring across a wider range of legal and economic questions.

**References**


Armacost, Barbara E. 2003. "Organizational Culture and Police Misconduct." *George Washington Law Review* 72 (3): 453–546.

Averett, Elliott. 2023. "An Unqualified Defense of Qualified Immunity Notes." *Georgetown Journal of Law & Public Policy* 21 (1): 241–77.

Balsamini, Dean. 2021. "Exclusive | Are NYPD Officers Rushing to Retire amid City's Anti-Cop Climate?" New York Post, April 24. https://nypost.com/2021/04/24/are-nypd-officers-rushing-to-retire-amid-citys-anti-cop-climate/.

Baude, William. 2018. "Is Qualified Immunity Unlawful." *California Law Review* 106 (1): 45–90.

Bidwell, Anya. 2023. "The Nevada Supreme Court Recognizes a Damages Remedy Directly Under Its State Constitution." The Federalist Society, June 7. https://fedsoc.org/commentary/publications/the-nevada-supreme-court-recognizes-a-damages-remedy-directly-under-its-state-constitution.

Bidwell, Anya, Patrick Jaicomo, and Megan Cairns. 2022a. "50 Shades of Government Immunity - California." Institute for Justice, January 25. https://ij.org/report/50-shades-of-government-immunity/state-profile/california/.

Bidwell, Anya, Patrick Jaicomo, and Megan Cairns. 2022b. "50 Shades of Government Immunity - Maryland." Institute for Justice, January 25. https://ij.org/report/50-shades-of-government-immunity/state-profile/maryland/.

Bidwell, Anya, Patrick Jaicomo, and Megan Cairns. 2022c. "50 Shades of Government Immunity - Michigan." Institute for Justice, January 25. https://ij.org/report/50-shades-of-government-immunity/state-profile/maryland/.

Carlos, Hayden. 2018. "Disqualifying Immunity: How Qualified Immunity Exacerbates Police Misconduct and Why Congress Must Destroy It." *Southern University Law Review* 46 (2): 283–324.

Chalfin, Aaron, and Justin McCrary. 2018. "Are U.S. Cities Underpoliced? Theory and Evidence." *The Review of Economics and Statistics* 100 (1): 167–86. https://doi.org/10.1162/REST_a_00694.

Chapman, Nathan S. 2023. "Fair Notice, the Rule of Law, and Reforming Qualified Immunity." *Florida Law Review* 75 (1): 1–64.

Crocker, Katherine Mims. 2021. "Qualified Immunity, Sovereign Immunity, and Systemic Reform." *Duke Law Journal* 71 (8): 1701–80.

Drysdale, Sam. 2025. “Police Oppose New Plan to Overhaul Qualified Immunity.” *The New Bedford Light*, July 31. https://newbedfordlight.org/police-oppose-new-plan-to-overhaul-qualified-immunity/.

Enhance Law Enforcement Integrity Act, SB20-217, Colorado General Assembly (2020). https://leg.colorado.gov/bills/sb20-217.

Fisher, Austin. 2025. “NM Municipal League Claims NM Civil Rights Act a ‘Complete Disaster’ for Police Recruitment • Source New Mexico.” *Source New Mexico*, July 1. https://sourcenm.com/2025/07/01/nm-municipal-league-claims-nm-civil-rights-act-a-complete-disaster-for-police-recruitment/.

“Frequently Asked Questions About Ending Qualified Immunity.” n.d. *Institute for Justice*. Accessed January 14, 2026. https://ij.org/issues/project-on-immunity-and-accountability/frequently-asked-questions-about-ending-qualified-immunity/.

Gelman, Andrew, and Eric Loken. 2014. “The Statistical Crisis in Science.” *American Scientist* 102 (6): 460. https://doi.org/10.1511/2014.111.460.

Gildin, Gary. 2024. “Legislative Efforts to Abolish Qualified Immunity Yield Mixed Results.” State Court Report, May 30. https://statecourtreport.org/our-work/analysis-opinion/legislative-efforts-abolish-qualified-immunity-yield-mixed-results.

Gilles, Myriam E. 2000. “In Defense of Making Government Pay: The Deterrent Effect of Constitutional Tort Remedies Symposium: Re-Examining First Principles: Deterrence and Corrective Justice in Constitutional Torts.” *Georgia Law Review* 35 (3): 845–80.

Karakatsanis, Alec. 2025. *Copaganda: How Police and the Media Manipulate Our News*. The New Press.

Karlan, Pamela S. 2006. “The Paradoxical Structure of Constitutional Litigation The Robert L. Levine Distinguished Lecture Series.” *Fordham Law Review* 75 (4): 1913–30.

Keene, Houston. 2023. “Banks Reintroduces Bill to Codify Qualified Immunity Police Protections into Law.” Text.Article. Fox News, Fox News, January 10. https://www.foxnews.com/politics/banks-reintroduces-bill-codify-qualified-immunity-police-protections-law.

Keyes, Vance D., and Latocia Keyes. 2023. “A Scoping Review of Police Civilian Oversight Literature 1992–2022.” *Policing: An International Journal* 46 (4): 601–22. https://doi.org/10.1108/PIJPSM-02-2023-0027.

Kim, Taeho. 2025. “Facilitating Police Reform: Body Cameras, Police-Involved Homicides, and Law Enforcement Outcomes.” *Journal of Public Economics* 248 (August): 105424. https://doi.org/10.1016/j.jpubeco.2025.105424.

Kindy, Kimberly. 2021. “Dozens of States Have Tried to End Qualified Immunity. Police Officers and Unions Helped Beat Nearly Every Bill.” *The Washington Post*, October 6.

https://www.washingtonpost.com/politics/qualified-immunity-police-lobbying-state-legislatures/2021/10/06/60e546bc-0cdf-11ec-aea1-42a8138f132a_story.html.

Leamer, Edward E. 1983. "Let's Take the Con Out of Econometrics." *The American Economic Review* 73 (1): 31–43.

Levine, Ross, and David Renelt. 1992. "A Sensitivity Analysis of Cross-Country Growth Regressions." *The American Economic Review* 82 (4): 942–63.

Levinson, Daryl J. 2000. "Making Government Pay: Markets, Politics, and the Allocation of Constitutional Costs." *University of Chicago Law Review* 67 (2): 345–420.

McDonald, Mark. 2021. "Police Recruitment Was Already Tough. Attacks on Qualified Immunity Make Matters Worse." *Fraternal Order of Police*, November 30. https://fop.net/2021/11/police-recruitment-was-already-tough-attacks-on-qualified-immunity-make-matters-worse/.

Merrell, Sydney. 2021. "Get out of Jail Free Card: Doctrine of Qualified Immunity." *Thurgood Marshall Law Review* 46 (1): 95–114.

New Mexico Civil Rights Act, New Mexico Statutes (2021). https://law.justia.com/codes/new-mexico/chapter-41/article-4a/.

Nielson, Aaron L., and Christopher J. Walker. 2017. "A Qualified Defense of Qualified Immunity Federal Courts, Practice & Procedure: The Future of Qualified Immunity." *Notre Dame Law Review* 93 (5): 1853–86.

Nielson, Aaron L., and Christopher J. Walker. 2020. "Qualified Immunity and Federalism." *Georgetown Law Journal* 109 (2): 229–304.

"Qualified Immunity." n.d. *Equal Justice Initiative*. Accessed January 14, 2026. https://eji.org/issues/qualified-immunity/.

Reinert, Alexander A. 2023. "Qualified Immunity's Flawed Foundation." *California Law Review* 111 (1): 201–46.

Rocha Beardall, Theresa. 2022. "Police Legitimacy Regimes and the Suppression of Citizen Oversight in Response to Police Violence." *Criminology* 60 (4): 740–65. https://doi.org/10.1111/1745-9125.12321.

Rosen, Jeffrey. 2020. "Sorry: This Criminal-Justice 'Reform' Would Be Disaster for Policing." New York Post, September 24. https://nypost.com/2020/09/24/sorry-this-criminal-justice-reform-would-be-disaster-for-policing/.

Rosenthal, Lawrence. 2020. "Defending Qualified Immunity." *South Carolina Law Review* 72 (2): [i]-616.

Rossler, Michael T., and Charles Scheer. 2024. "Causes of Police Officer Career Apprehension Following George Floyd." *Police Quarterly* 27 (4): 505–31. https://doi.org/10.1177/10986111241234317.

Sala-I-Martin, Xavier, Gernot Doppelhofer, and Ronald I. Miller. 2004. "Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach." *American Economic Review* 94 (4): 813–35. https://doi.org/10.1257/0002828042002570.

Sala-I-Martin, Xavier X. 1997. "I Just Ran Two Million Regressions." *The American Economic Review* 87 (2): 178–83.

Santos, Mateus R., Chae M. Jaynes, and Danielle M. Thomas. 2024. "Informing the Recruitment Crisis in Policing: Evaluating Which Incentives Can Entice the Best Candidates among College Students." *Criminology & Public Policy* 23 (3): 569–603. https://doi.org/10.1111/1745-9133.12668.

Schwartz, Joanna. 2023. *Shielded: How the Police Became Untouchable*. Viking.

Schwartz, Joanna C. 2009. "Myths and Mechanics of Deterrence: The Role of Lawsuits in Law Enforcement Decisionmaking." *UCLA Law Review* 57 (4): 1023–94.

Schwartz, Joanna C. 2014. "Police Indemnification." *New York University Law Review* 89 (3): 885–1005.

Schwartz, Joanna C. 2017a. "How Qualified Immunity Fails." *Yale Law Journal* 127 (1): 2–77.

Schwartz, Joanna C. 2017b. "The Case against Qualified Immunity." *Notre Dame Law Review* 93 (5): 1797–852.

Schwartz, Joanna C. 2019. "Qualified Immunity's Selection Effects." *Northwestern University Law Review* 114 (5): 1101–78.

Schweikert, Jay. 2020. "Qualified Immunity: A Legal, Practical, and Moral Failure." Cato Institute, September 14. https://www.cato.org/policy-analysis/qualified-immunity-legal-practical-moral-failure.

Sherry, Allison. 2022. "After Police and Sheriffs Deputies Left Agencies in Droves in 2021, Democratic Leaders Try to Stem the Tide." Colorado Public Radio, January 13. https://www.cpr.org/2022/01/13/colorado-law-enforcement-hiring-attrition/.

Sibilla, Nick. 2020. "New Connecticut Law Limits Police Immunity In Civil Rights Lawsuits, But Loopholes Remain." Forbes, July 31. https://www.forbes.com/sites/nicksibilla/2020/07/31/new-connecticut-law-limits-police-immunity-in-civil-rights-lawsuits-but-loopholes-remain/.

Sibilla, Nick. 2021. "New York City Bans Qualified Immunity For Cops Who Use Excessive Force." Forbes, April 30. https://www.forbes.com/sites/nicksibilla/2021/04/29/new-york-city-limits-qualified-immunity-makes-it-easier-to-sue-cops-who-use-excessive-force/.

Simmons, Joseph P., Leif D. Nelson, and Uri Simonsohn. 2011. “False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant.” *Psychological Science* 22 (11): 1359–66. https://doi.org/10.1177/0956797611417632.

Widra, Wendy Sawyer, Emily. n.d. “Resource Spotlight: Data Projects Tracking Police Misconduct, Use of Force, and Employment Histories.” Accessed January 27, 2026. https://www.prisonpolicy.org/blog/2026/01/26/police_misconduct/.

Wolf, Barry. 2020. *Guidance on Using the Bane Act to Redress Government Misconduct*. 43 (March): 10–12.

Zare, Hossein, Danielle R. Gilmore, Khushbu Balsara, et al. 2025. “State-by-State Review: The Spread of Law Enforcement Accountability Policies.” *Social Sciences* 14 (8): 483. https://doi.org/10.3390/socsci14080483.

**Tables and Figures**

**Figure 1: Trends Over Time in Police Employment in Colorado Versus Comparison Groups**

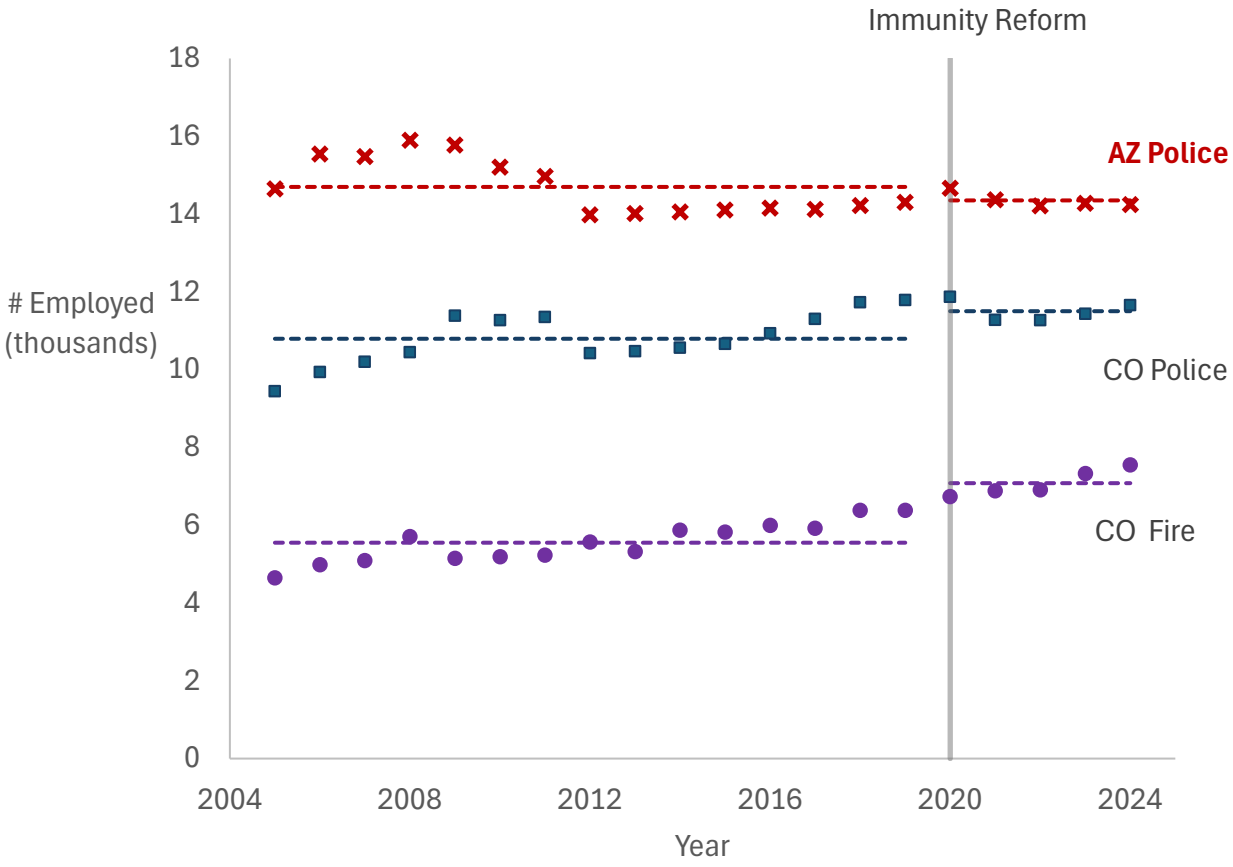
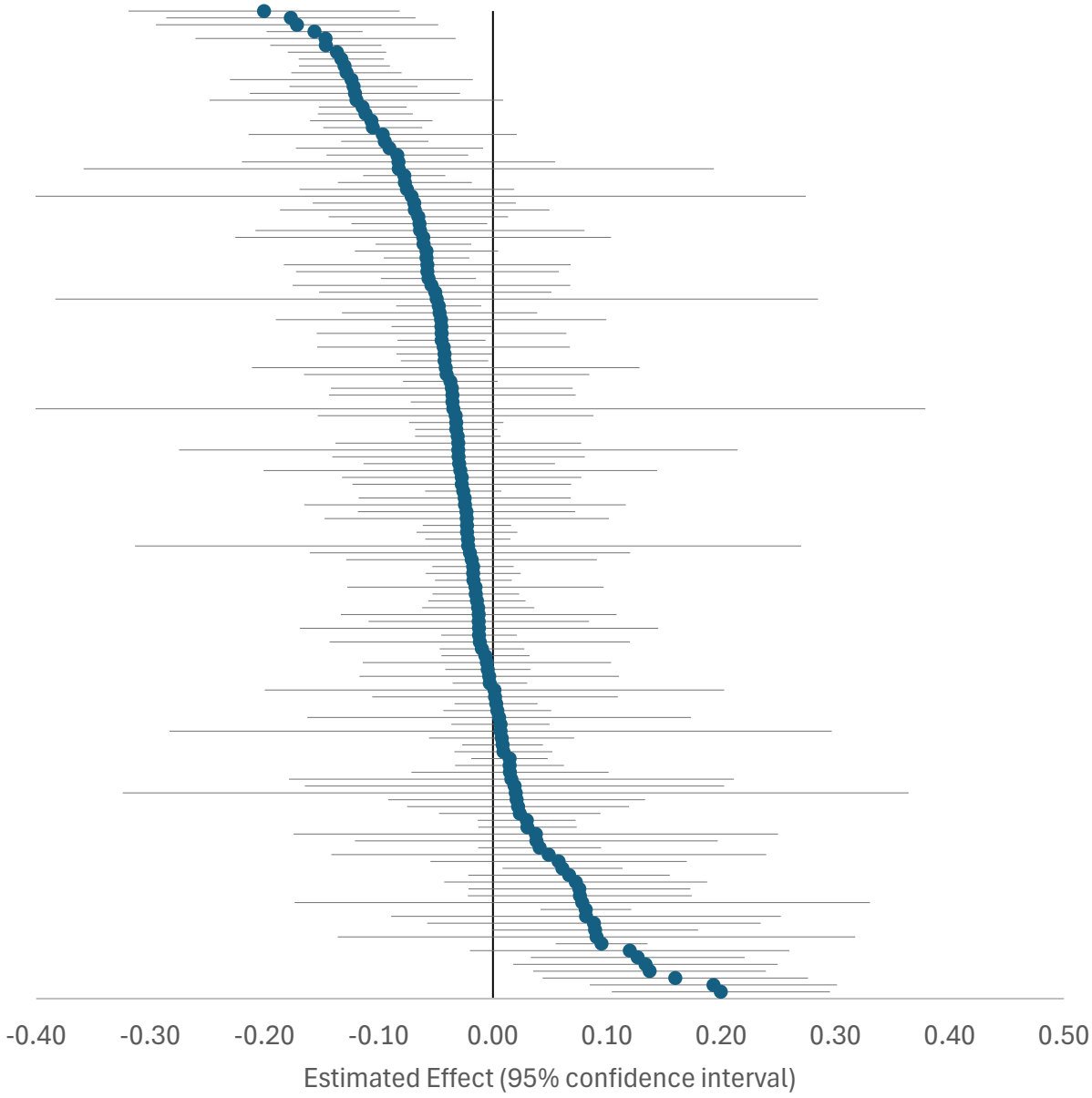
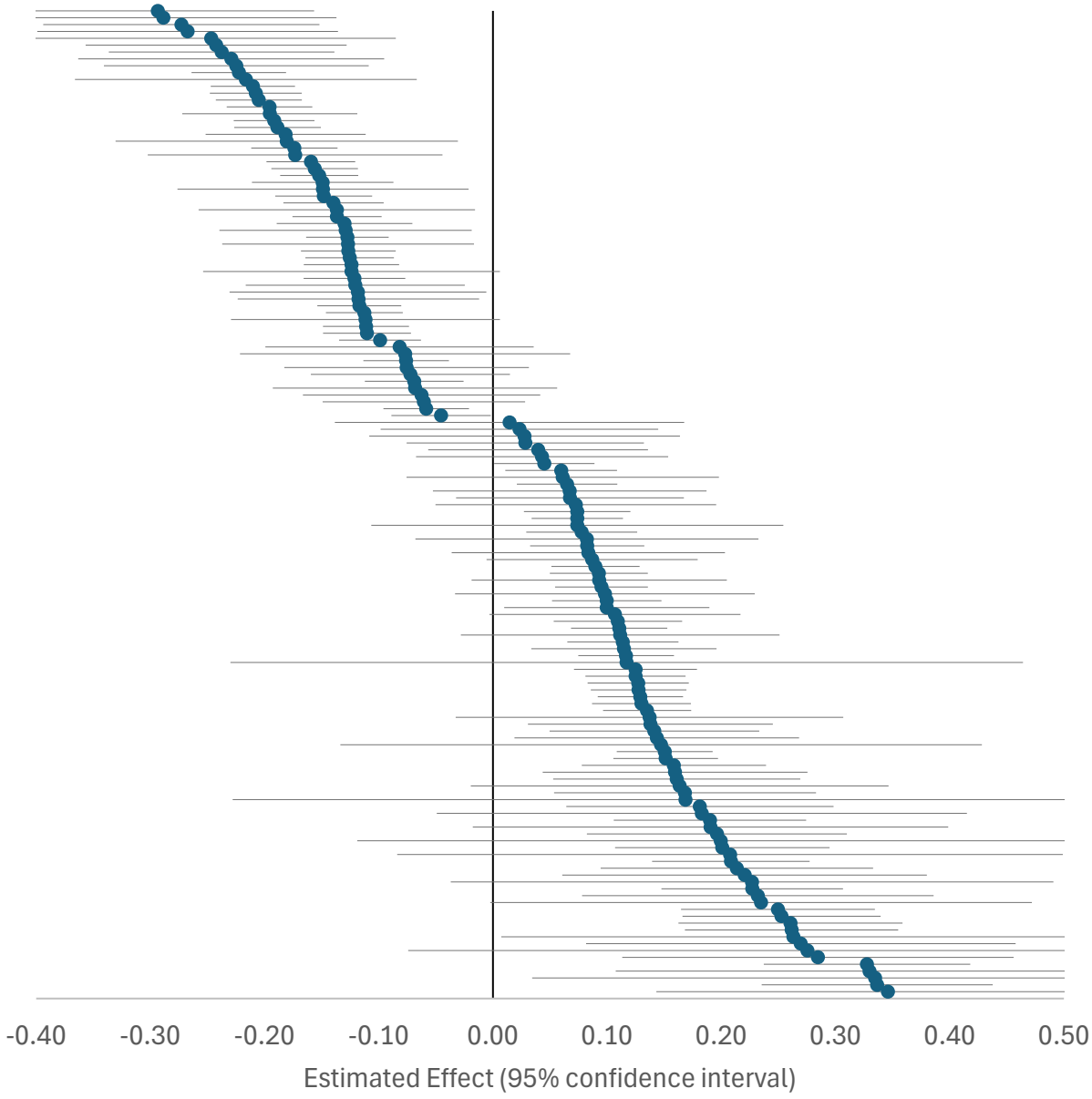


**Figure 2: Estimated Effects of QI Reform on Log(Police Employment) Across Specifications, American Community Survey (ACS)**

**a. Colorado**



**b. New Mexico**


-0.40
-0.30
-0.20
-0.10
0.00
0.10
0.20
0.30
0.40
0.50
Estimated Effect (95% confidence interval)

**c. Nevada**

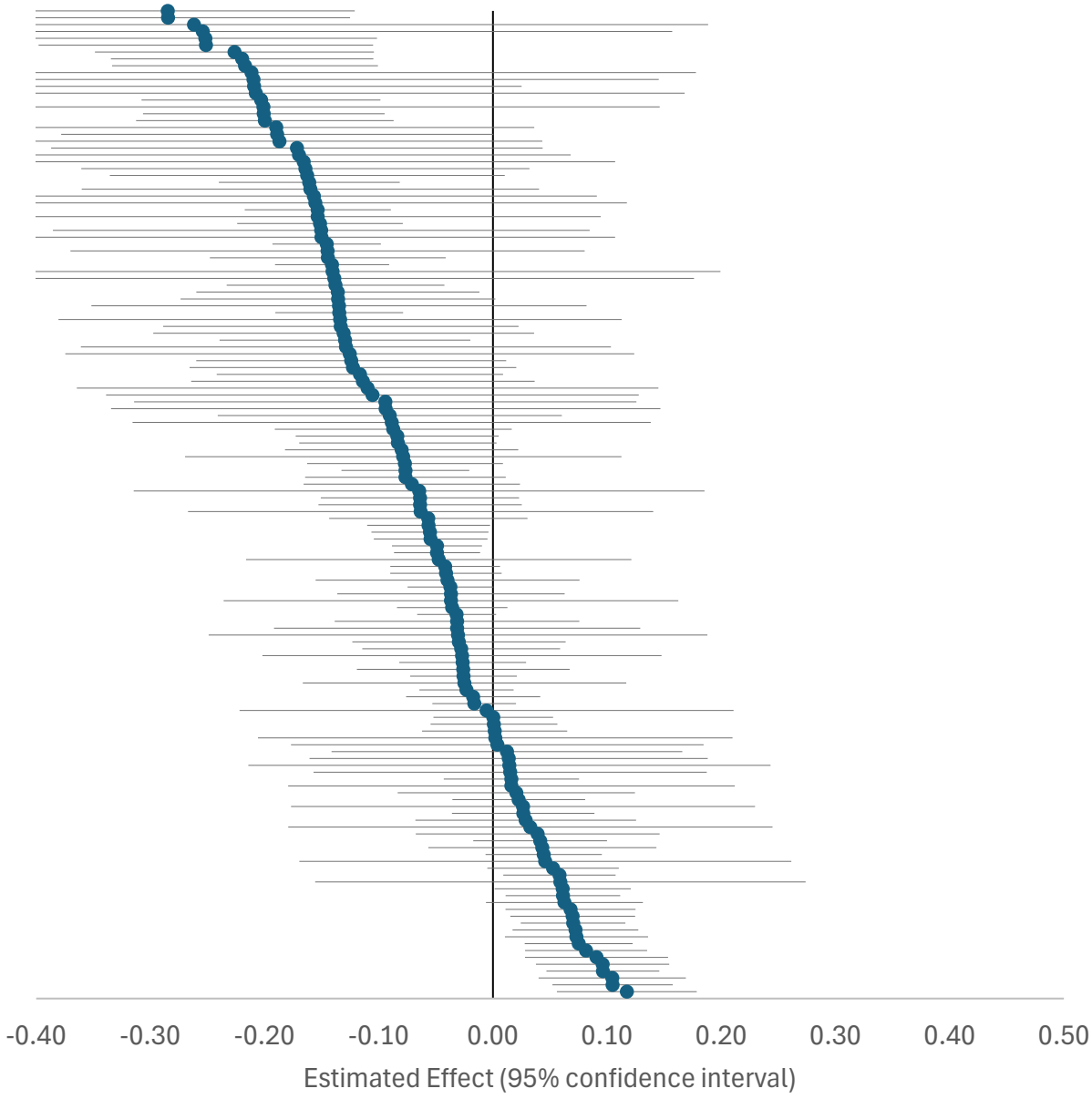
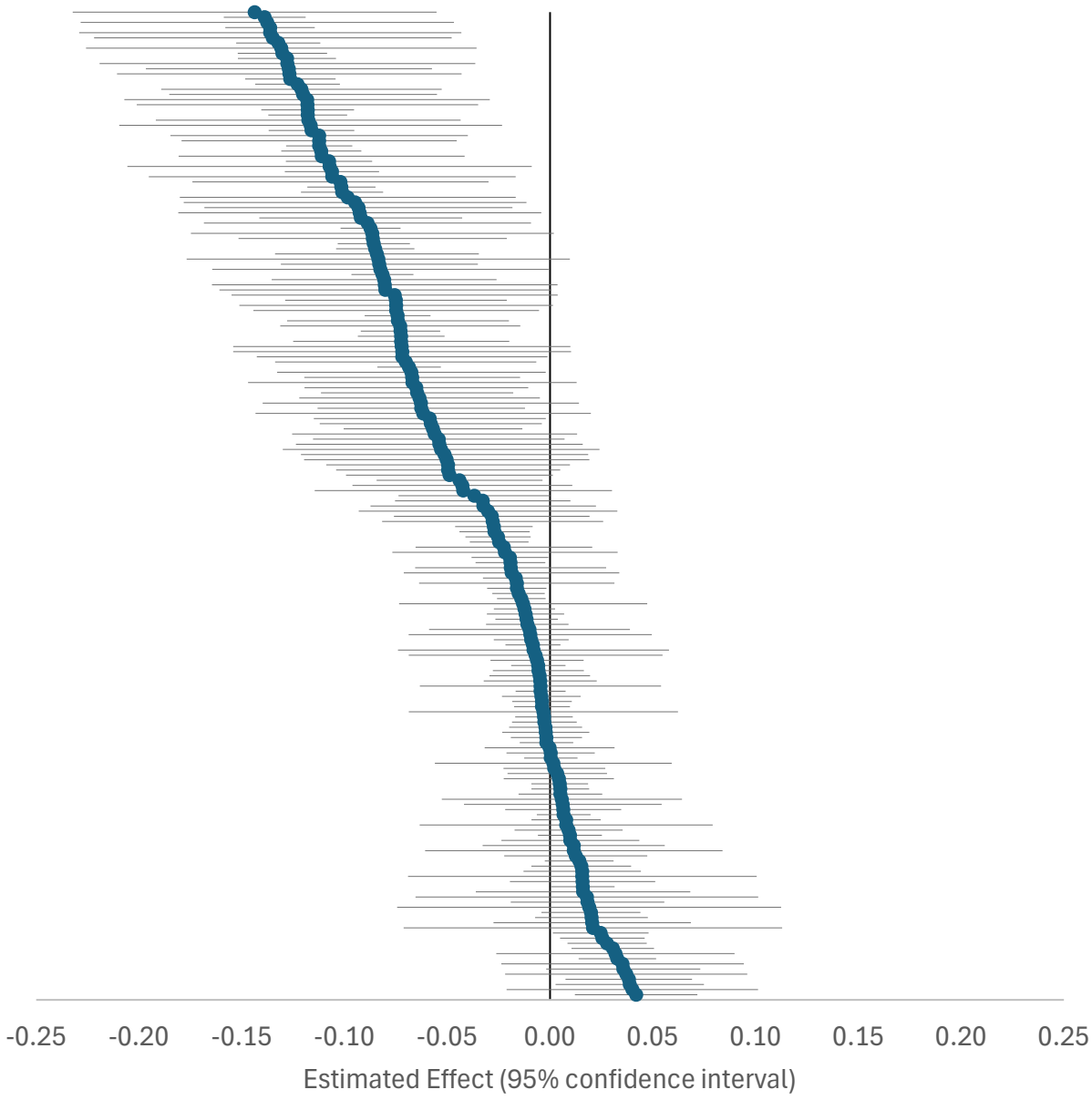
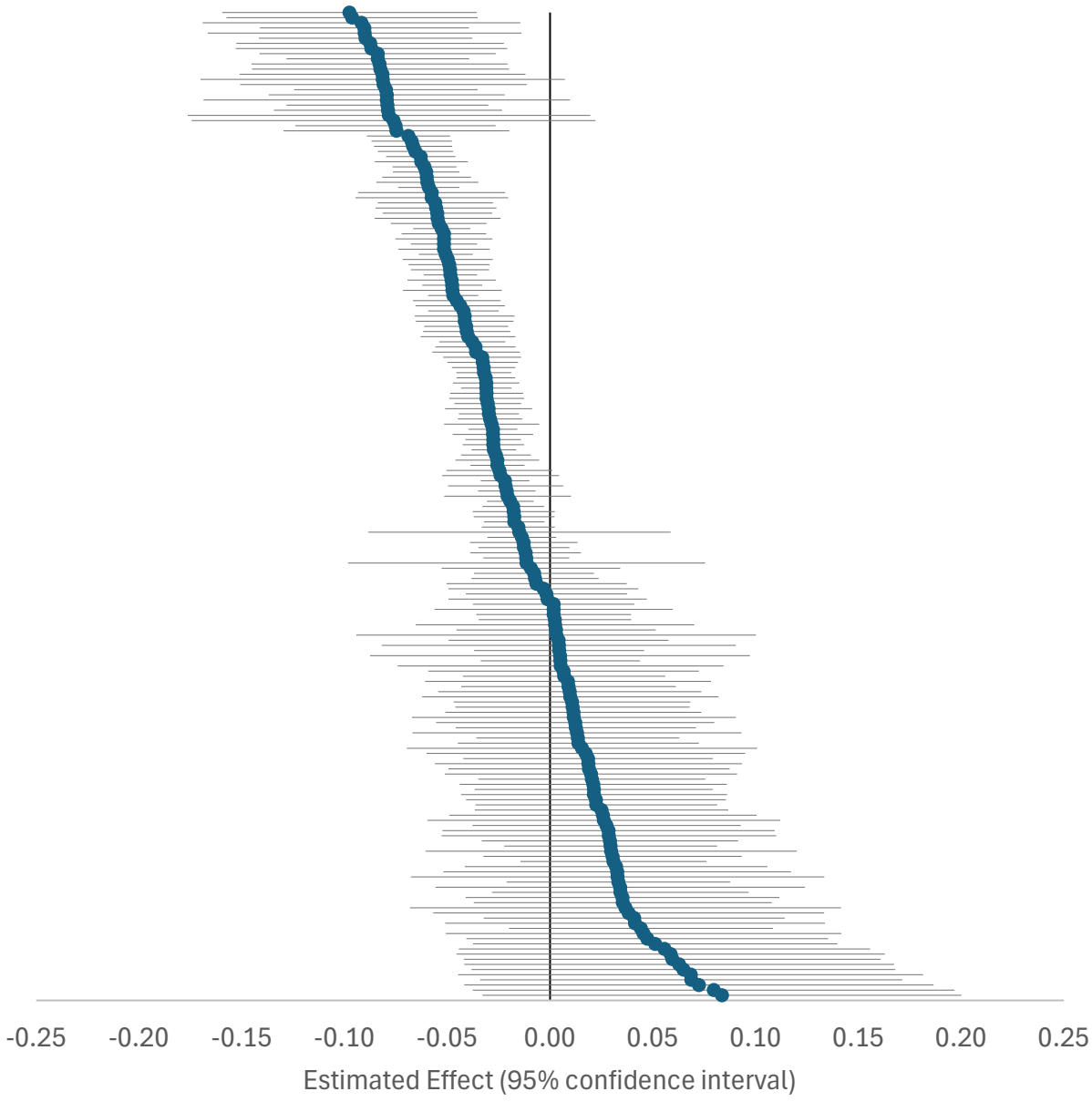

-0.40
-0.30
-0.20
-0.10
0.00
0.10
0.20
0.30
0.40
0.50
Estimated Effect (95% confidence interval)

**Figure 3: Estimated Effects of QI Reform on Log(Police Employment) Across Specifications, Annual Survey of Public Employment and Payroll (ASPEP)**

**a. Colorado**



**b. New Mexico**


-0.25
-0.20
-0.15
-0.10
-0.05
0.00
0.05
0.10
0.15
0.20
0.25
Estimated Effect (95% confidence interval)

**c. Nevada**

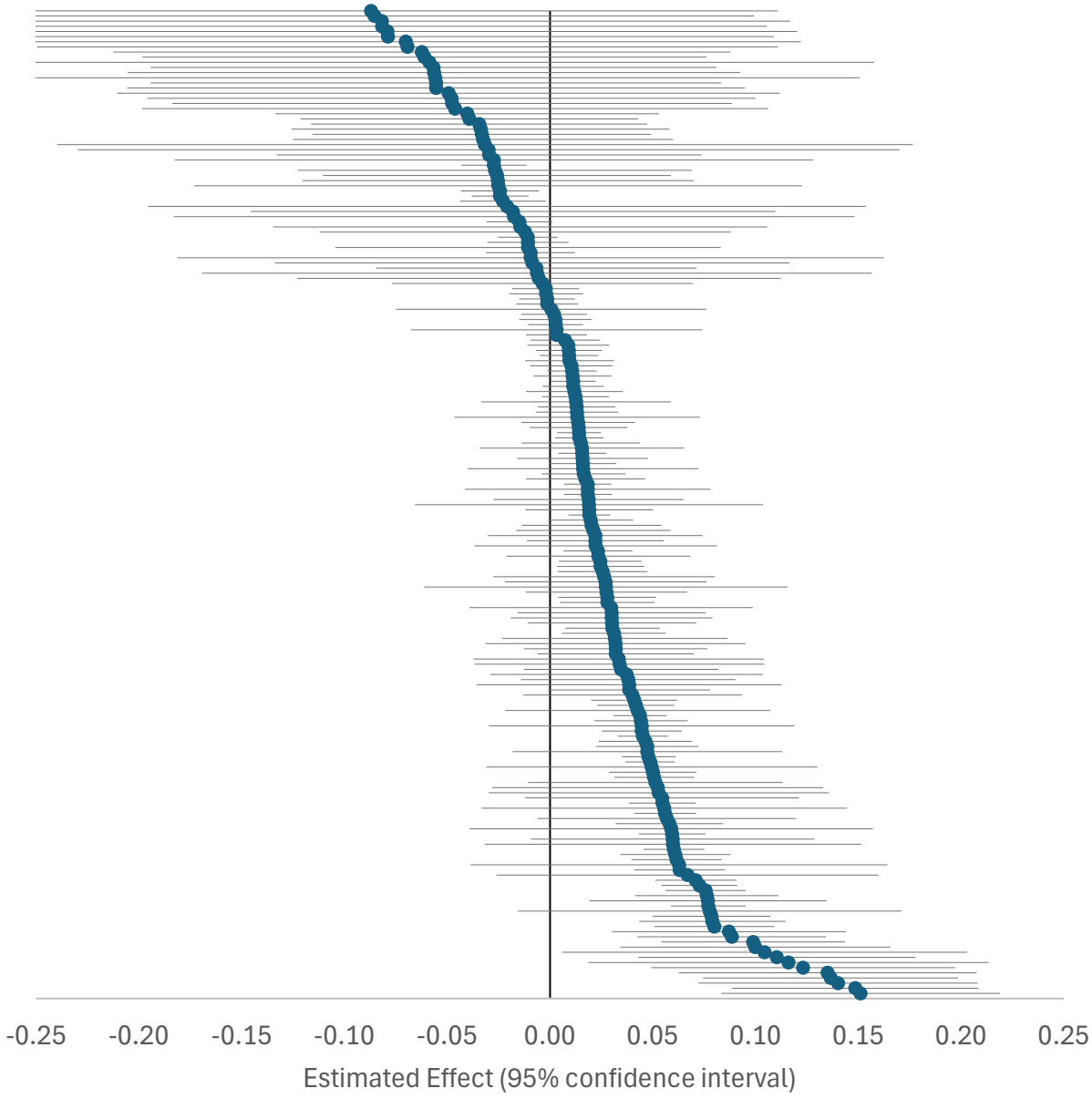
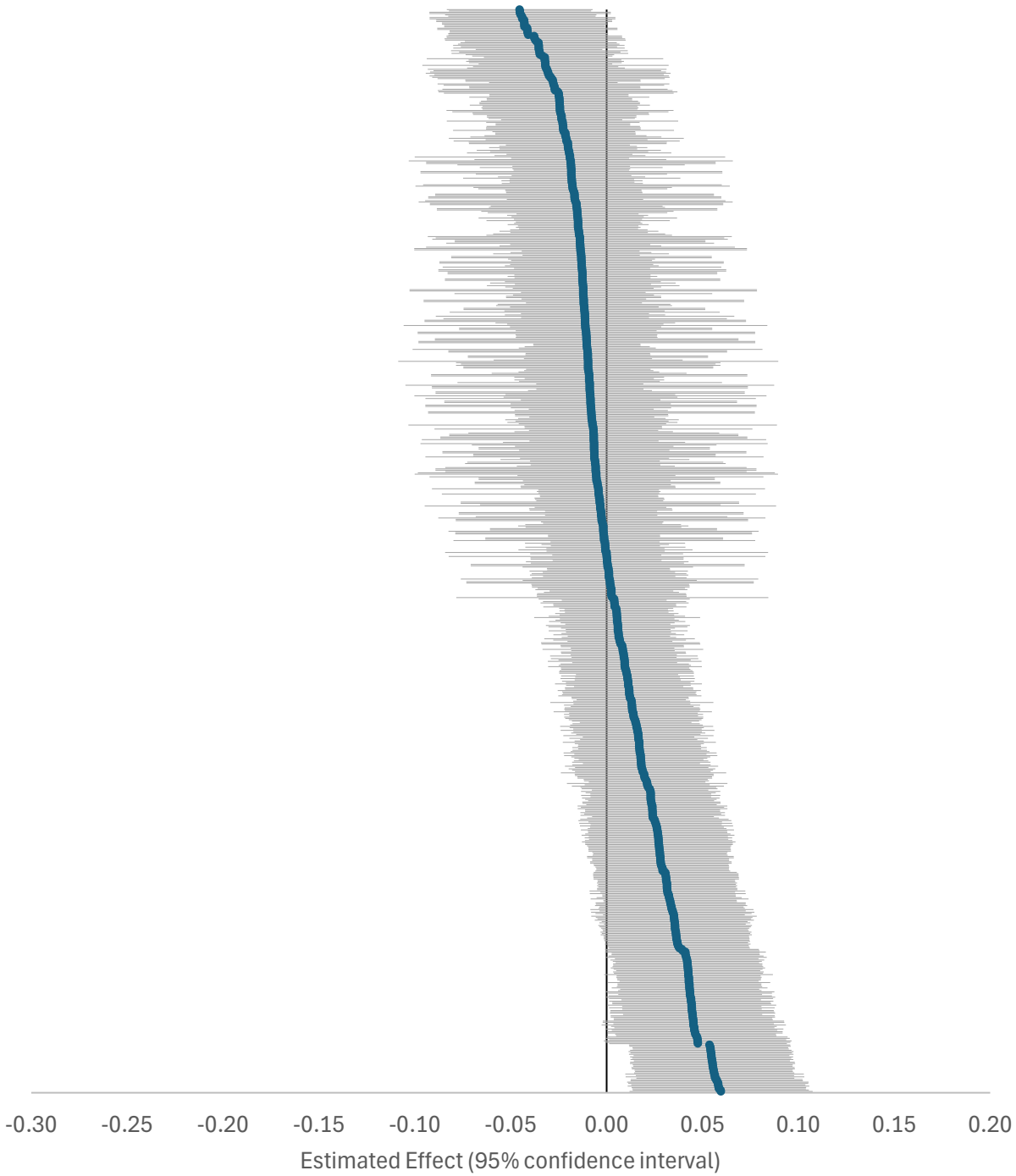
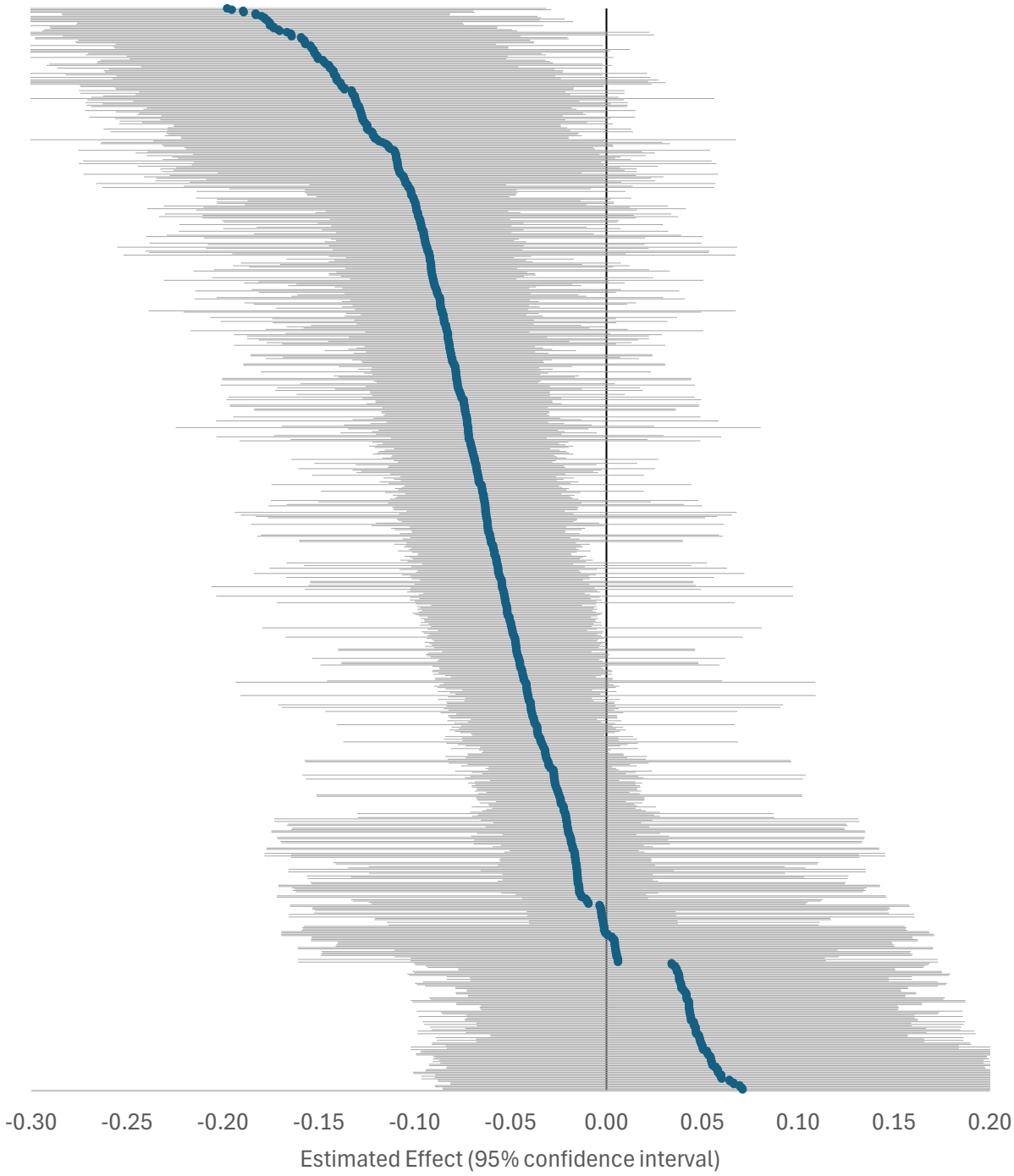

-0.25
-0.20
-0.15
-0.10
-0.05
0.00
0.05
0.10
0.15
0.20
0.25
Estimated Effect (95% confidence interval)

**Figure 4: Estimated Effects of QI Reform on Log(Police Employment) Across Specifications, Law Enforcement Employment (LEE)**

**a. Colorado**



**b. New Mexico**


-0.30
-0.25
-0.20
-0.15
-0.10
-0.05
0.00
0.05
0.10
0.15
0.20
Estimated Effect (95% confidence interval)

**c. Nevada**

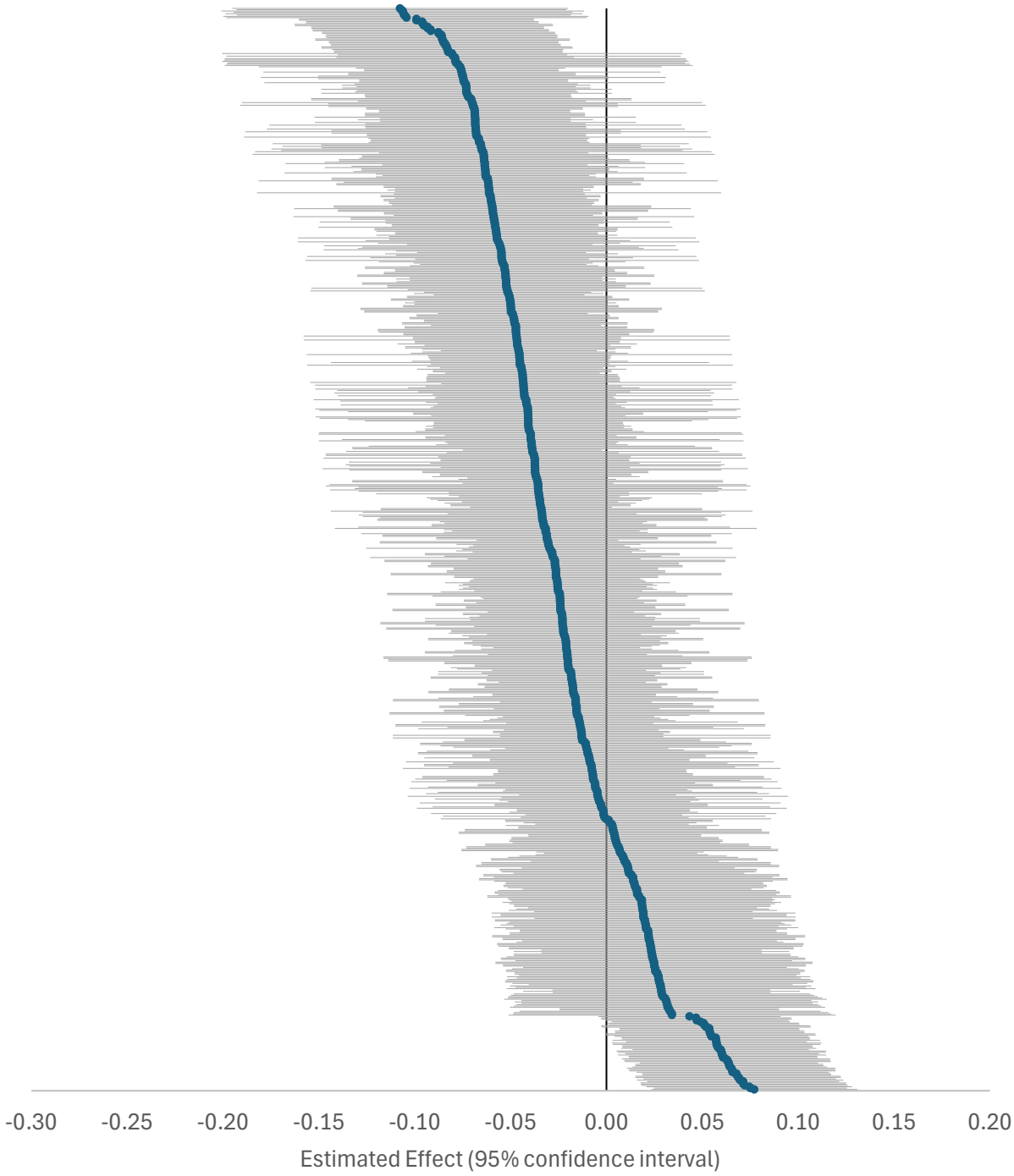

-0.30
-0.25
-0.20
-0.15
-0.10
-0.05
0.00
0.05
0.10
0.15
0.20
Estimated Effect (95% confidence interval)

**Figure 5: Combined Distributions**

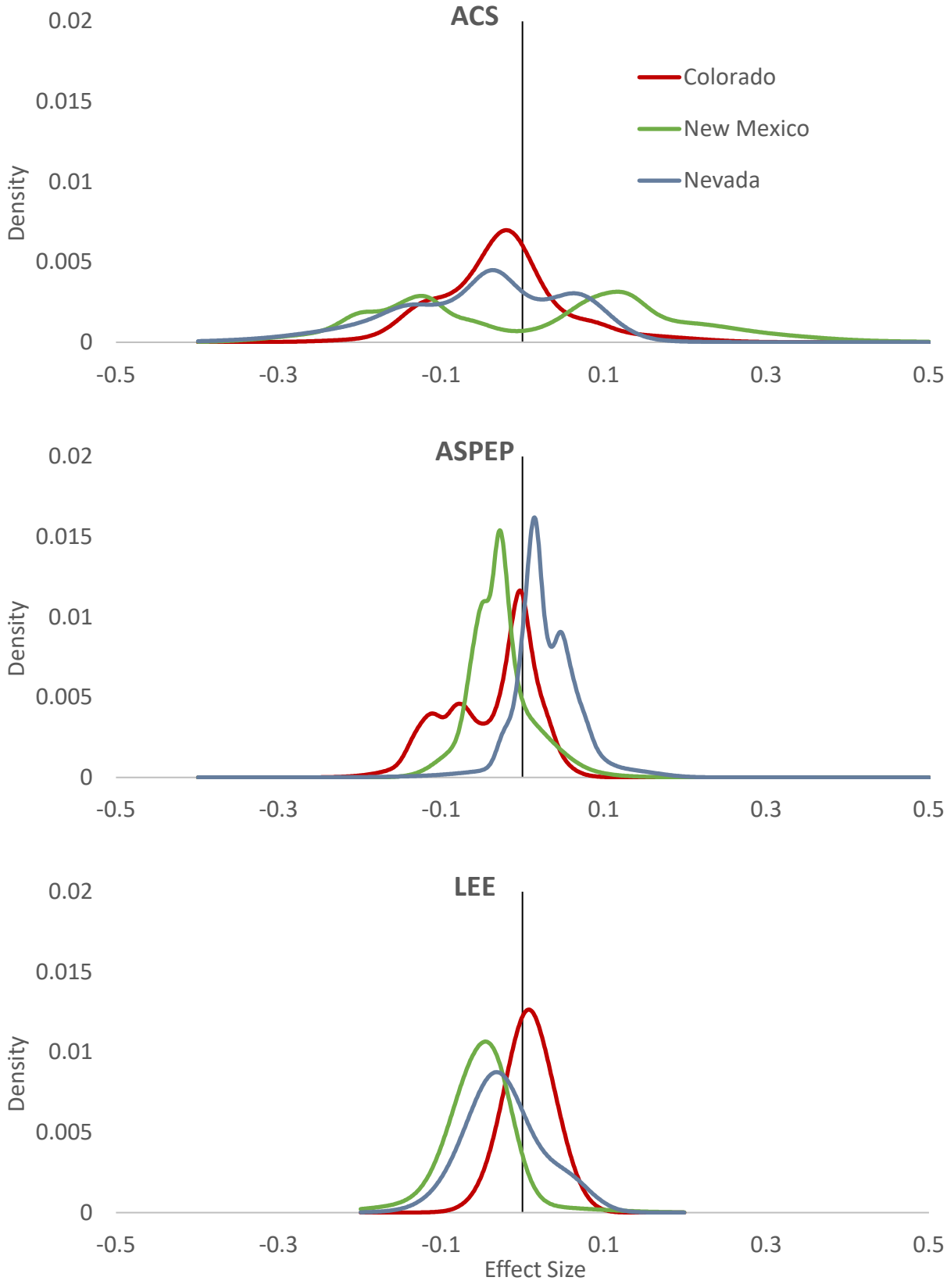


**Table 1: Differences-in-Differences (DD) and Triple Differences (DDD) Estimates of the Effects of QI Reform, Annual Survey of Public Employment and Payroll (ASPEP)**

| | I | II |
|---|---|---|
| **Estimated Effect For:** | **DD** | **DDD** |
| Colorado | .027** | -.135** |
| | (.010) | (.044) |
| New Mexico | -.067*** | .014 |
| | (.009) | (.030) |
| Nevada | .072*** | -.079 |
| | (.009) | (.101) |
| N | 1018 | 5089 |
| $R^2$ | .997 | .995 |
| Include state and year fixed effects? | Yes | Yes |
| Include state/year, industry/year, and industry state fixed effects? | No | Yes |

Note: This table reports results from DD and DDD regressions measuring the effect of qualified immunity reform on log(police employment). The unit of observation for the DD regressions is a state/year, and for the DDD regressions is a state/industry/year. The primary explanatory variables of interest in the DD regressions are indicators for Colorado on or after 2020; New Mexico on or after 2021; and Nevada on or after 2023. In the DDD regressions we further interact these with indicators for police. The sample includes all U.S. states and DC from 2005-2024. Control industries include counts of firefighters, highway workers, public health workers, and health workers who are government employees. Each column reports results from a separate regression. Standard errors clustered on state (DD) or state/industry (DDD) are reported in parentheses. * denotes a statistically significant estimate at the 5% level; ** at the 1% level, and *** at the .1% level.

**Table 2: Alternative Specifications Considered in Analysis**

**a. ACS**

| Analysis Dimension | Variant 1 | Variant 2 | Variant 3 |
|---|---|---|---|
| Weighting | Every state counts equally | Weight by state population | |
| Control Group | All states | Neighboring states | |
| State Exclusions | All states | Exclude states with other immunity reforms | |
| Comparison Industry | Private security | Firefighters | |
| Sample Start Date | 2005 | 2010 | 2015 |
| Identification Approach | Differences-in-differences | Triple differences | |
| Treatment Definition | Include first year of implementation | Exclude first year of implementation | |

**b. ASPEP**

| Analysis Dimension | Variant 1 | Variant 2 | Variant 3 |
|---|---|---|---|
| Dependent Variables | Levels | Per capita | |
| Weighting | Every state counts equally | Weight by state population | |
| Control Group | All states | Neighboring states | |
| State Exclusions | All states | Exclude states with other immunity reforms | |
| Comparison Industry | Firefighters | Firefighters + four other industries | |
| Sample Start Date | 2005 | 2010 | 2015 |
| Identification Approach | Differences-in-differences | Triple differences | |
| Treatment Definition | Include first year of implementation | Exclude first year of implementation | |

### c. LEE

| Analysis Dimension | Variant 1 | Variant 2 | Variant 3 |
|---|---|---|---|
| Weighting | Every agency counts equally | Weight by agency size | |
| Control Group | All states | Neighboring states | |
| State Exclusions | All states | Exclude states with other immunity reforms | |
| Sample Composition | All agencies | Agencies with at least 10 officers | |
| Sample Start Date | 2003 | 2010 | 2015 |
| Data Availability | No restriction | Agency must report at least 11 years | Agency must report at least 20 years |
| Data Quality | No restriction | Exclude state/years with <50% of agencies reporting | Agency data shows no "jumps" |
| Treatment Definition | Include first year of implementation | Exclude first year of implementation | |

**Table 3: Midpoints and 95% Confidence Intervals for Combined Distributions**

| Dataset | Colorado | New Mexico | Nevada |
|---|---|---|---|
| ACS | -.027 | .042 | -.041 |
| | [-.171, .161] | [-.275, .343] | [-.293, .125] |
| ASPEP | -.022 | -.031 | .023 |
| | [-.149, .046] | [-.101, .067] | [-.046, .121] |
| LEE | .007 | -.052 | -.027 |
| | [-.057, .071] | [-.156, .039] | [-.117, .084] |

**Table 4: Relationship Between Modeling Design Decisions and Model Estimates**

**a. Estimates**

| | ACS | | | ASPEP | | | LEE | | |
|---|---|---|---|---|---|---|---|---|---|
| **Design Decision** | **CO** | **NM** | **NV** | **CO** | **NM** | **NV** | **CO** | **NM** | **NV** |
| Dependent variable | | | | | | | | | |
| Reference: Measure in levels | | | | | | | | | |
| Measure in per capita rates | | | | -0.012** | 0.039** | -0.014* | | | |
| | | | | (0.003) | (0.004) | (0.006) | | | |
| Population Weight | | | | | | | | | |
| Reference: No | | | | | | | | | |
| Yes | 0.039** | 0.041** | 0.069** | 0.010** | 0.009** | 0.008 | 0.019** | 0.045** | 0.034** |
| | (0.008) | (0.013) | (0.010) | (0.003) | (0.003) | (0.005) | (0.001) | (0.001) | (0.002) |
| Treatment Begins | | | | | | | | | |
| Reference: Year of enactment | | | | | | | | | |
| Year after enactment | -0.065** | -0.078** | 0.061** | 0.011** | -0.005 | 0.001 | 0.003** | 0.021** | -0.012** |
| | (0.008) | (0.014) | (0.009) | (0.003) | (0.003) | (0.005) | (0.001) | (0.001) | (0.002) |
| State Sample | | | | | | | | | |
| Reference: All states | | | | | | | | | |
| Census Mountain division | 0.034** | 0.028* | -0.062** | 0.018** | 0.020** | 0.025** | -0.036** | -0.040** | -0.032** |
| | (0.008) | (0.014) | (0.010) | (0.003) | (0.004) | (0.005) | (0.001) | (0.001) | (0.002) |
| State Exclusions | | | | | | | | | |
| Reference: None | | | | | | | | | |
| Exclude states with other reforms | -0.022** | -0.020 | -0.012 | 0.002 | 0.006 | 0.002 | 0.004** | 0.004** | 0.004** |
| | (0.008) | (0.014) | (0.010) | (0.003) | (0.003) | (0.005) | (0.001) | (0.001) | (0.002) |
| Years Included | | | | | | | | | |
| Reference: 2005-present | | | | | | | | | |
| 2010-present | 0.008 | -0.017 | 0.017 | -0.003 | -0.002 | -0.009 | -0.005** | -0.000 | -0.014** |
| | (0.009) | (0.014) | (0.012) | (0.004) | (0.004) | (0.007) | (0.001) | (0.002) | (0.002) |
| 2015-present | 0.047** | -0.110** | 0.029* | 0.003 | 0.001 | -0.023** | -0.014** | 0.014** | -0.045** |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | (0.010) | (0.016) | (0.012) | (0.004) | (0.004) | (0.007) | (0.001) | (0.002) | (0.002) |
| Diffs-in-diffs approach | | | | | | | | | |
| Reference: DDD | | | | | | | | | |
| DD | -0.005 | -0.273** | 0.069** | 0.088** | -0.042** | 0.026** | | | |
| | (0.009) | (0.013) | (0.012) | (0.003) | (0.003) | (0.005) | | | |
| Comparison industries | | | | | | | | | |
| Reference: Firefighters | | | | | | | | | |
| Private security | -0.040** | 0.008 | 0.100** | | | | | | |
| | (0.008) | (0.014) | (0.010) | | | | | | |
| Firefighters + 4 other industries | | | | -0.004 | 0.020** | -0.042** | | | |
| | | | | (0.003) | (0.004) | (0.005) | | | |
| Sample Composition | | | | | | | | | |
| Reference: All agencies | | | | | | | | | |
| Agencies with at least 10 officers | | | | | | | -0.009** | -0.017** | 0.001 |
| | | | | | | | (0.001) | (0.001) | (0.002) |
| Data Availability | | | | | | | | | |
| Reference: All agencies | | | | | | | | | |
| Reports in >11 years | | | | | | | -0.004** | -0.009** | 0.018** |
| | | | | | | | (0.001) | (0.002) | (0.002) |
| Reports in >20 years | | | | | | | -0.010** | -0.078** | 0.027** |
| | | | | | | | (0.001) | (0.002) | (0.002) |
| Data Quaility | | | | | | | | | |
| Reference: No restriction | | | | | | | | | |
| Exclude state/years with <50% | | | | | | | 0.000 | -0.002 | 0.000 |
| of agencies reporting | | | | | | | (0.001) | (0.002) | (0.002) |
| Agency data shows no jumps | | | | | | | 0.009** | -0.035** | 0.020** |
| | | | | | | | (0.001) | (0.002) | (0.002) |
| N | 144 | 144 | 144 | 192 | 192 | 192 | 816 | 816 | 816 |
| $R^2$ | 0.539 | 0.775 | 0.662 | 0.851 | 0.698 | 0.456 | 0.779 | 0.858 | 0.688 |

**b. Standard Errors**

| | ACS | | | ASPEP | | | LEE | | |
|---|---|---|---|---|---|---|---|---|---|
| **Design Decision** | **CO** | **NM** | **NV** | **CO** | **NM** | **NV** | **CO** | **NM** | **NV** |
| Dependent variable | | | | | | | | | |
| Reference: Measure in levels | | | | | | | | | |
| Measure in per capita rates | | | | -0.004** | -0.004** | -0.005 | | | |
| | | | | (0.001) | (0.001) | (0.002) | | | |
| Population Weight | | | | | | | | | |
| Reference: No | | | | | | | | | |
| Yes | -0.015** | -0.014** | -0.023** | 0.000 | 0.001 | -0.003 | 0.009** | 0.023** | 0.014** |
| | (0.003) | (0.003) | (0.003) | (0.001) | (0.001) | (0.002) | (0.000) | (0.001) | (0.000) |
| Treatment Begins | | | | | | | | | |
| Reference: Year of enactment | | | | | | | | | |
| Year after enactment | -0.006 | 0.002 | 0.009** | 0.001 | 0.002 | -0.001 | -0.001** | 0.001 | 0.001* |
| | (0.003) | (0.003) | (0.003) | (0.001) | (0.001) | (0.002) | (0.000) | (0.001) | (0.000) |
| State Sample | | | | | | | | | |
| Reference: All states | | | | | | | | | |
| Census Mountain division | 0.038** | 0.041** | 0.058** | 0.011** | 0.015** | 0.012** | 0.009** | 0.005** | 0.005** |
| | (0.003) | (0.003) | (0.003) | (0.001) | (0.001) | (0.002) | (0.000) | (0.001) | (0.000) |
| State Exclusions | | | | | | | | | |
| Reference: None | | | | | | | | | |
| Exclude states with other reforms | 0.001 | 0.002 | 0.005 | 0.000 | -0.000 | 0.000 | -0.001* | -0.000 | -0.000 |
| | (0.003) | (0.003) | (0.003) | (0.001) | (0.001) | (0.002) | (0.000) | (0.001) | (0.000) |
| Years Included | | | | | | | | | |
| Reference: 2005-present | | | | | | | | | |
| 2010-present | 0.006 | 0.004 | 0.002 | -0.006** | -0.005** | -0.011** | -0.003** | -0.008** | 0.003** |
| | (0.004) | (0.004) | (0.004) | (0.001) | (0.001) | (0.003) | (0.000) | (0.001) | (0.000) |
| 2015-present | -0.014** | -0.017** | -0.018** | -0.012** | -0.009** | -0.022** | -0.007** | -0.011** | 0.008** |
| | (0.004) | (0.004) | (0.004) | (0.001) | (0.001) | (0.003) | (0.000) | (0.001) | (0.001) |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Diffs-in-diffs approach | | | | | | | | | |
| Reference: DDD | | | | | | | | | |
| DD | -0.025** | -0.023** | -0.027** | -0.015** | -0.016** | -0.032** | | | |
| | (0.003) | (0.003) | (0.003) | (0.001) | (0.001) | (0.002) | | | |
| Comparison industries | | | | | | | | | |
| Reference: Firefighters | | | | | | | | | |
| Private security | -0.024** | -0.023** | -0.016** | | | | | | |
| | (0.003) | (0.003) | (0.003) | | | | | | |
| Firefighters + 4 other industries | | | | 0.009** | 0.010** | 0.026** | | | |
| | | | | (0.001) | (0.001) | (0.003) | | | |
| Sample Composition | | | | | | | | | |
| Reference: All agencies | | | | | | | | | |
| Agencies with at least 10 officers | | | | | | | 0.001** | 0.005** | -0.001** |
| | | | | | | | (0.000) | (0.001) | (0.000) |
| Data Availability | | | | | | | | | |
| Reference: All agencies | | | | | | | | | |
| Reports in >11 years | | | | | | | 0.000 | 0.001 | 0.003** |
| | | | | | | | (0.000) | (0.001) | (0.000) |
| Reports in >20 years | | | | | | | 0.000 | 0.023** | 0.005** |
| | | | | | | | (0.000) | (0.001) | (0.000) |
| Data Quaility | | | | | | | | | |
| Reference: No restriction | | | | | | | | | |
| Exclude state/years with <50% | | | | | | | 0.000 | -0.000 | 0.000 |
| of agencies reporting | | | | | | | (0.000) | (0.001) | (0.000) |
| Agency data shows no jumps | | | | | | | 0.000 | -0.013** | -0.002** |
| | | | | | | | (0.000) | (0.001) | (0.000) |
| N | 144 | 144 | 144 | 192 | 192 | 192 | 816 | 816 | 816 |
| $R^2$ | 0.671 | 0.697 | 0.775 | 0.698 | 0.741 | 0.643 | 0.748 | 0.622 | 0.677 |

**Table 5: Combined Meta-Analytic Estimates**

| State | I | II |
|---|---|---|
| Colorado | -.007 | -.010 |
| | (-.058, .044) | (-.061, .041) |
| New Mexico | -.037 | -.032 |
| | (-.094, .020) | (-.094, .030) |
| Nevada | .004 | .007 |
| | (-.055, .063) | (-.055, .068) |

Note: This table reports meta-analytic estimates that combine information from the three different datasets. 95% confidence intervals are reported in parentheses. Each table entry is derived from a separate meta-analysis. Column I reports estimates based on a normal approximation to the aggregated distributions shown in Figure 5. Column II reports estimates where the underlying "study" distributions have been chosen to match the confidence intervals reported in Table 3.

## Appendix

### Appendix Table 1: Estimates Based on Randomization Inference

| Dataset | Colorado | New Mexico | Nevada |
|---|---|---|---|
| ACS | | | |
| ASPEP | | | |
| LEE | | | |
| Combined | | | |

[replace with updated numbers]

## Other States with Qualified Immunity Reforms

In addition to Colorado, New Mexico, Nevada, and Montana, which have instituted fairly straightforward state workarounds to federal qualified immunity law, there are a number of states whose existing law is somewhere in between a full accommodation of federal qualified immunity law and creating a complete workaround. For each of these states, there is enough ambiguity in existing state law for a researcher to reasonably argue for or against their inclusion as control states. Consistent with our approach of remaining agnostic as to the "correct" specification, we ran some variants of our analysis with these states included as control states and others which excluded them.[24]

California

In 1987, California passed the Tom Bane Civil Rights Act (the Bane Act). This law allows individuals deprived of their federal and state constitutional rights to sue offending government officials when the offense is carried out by threat, intimidation, or coercion. California courts have not recognized qualified immunity as a defense to suits under the Bane Act (Bidwell et al. 2022a). Nonetheless, there is some disagreement among experts as to whether the Bane Act represents a strong work-around to §1983, with some arguing the standards for successfully suing offending officials are roughly similar across the statutes(Kindy 2021), and others viewing the Bane Act as offering plaintiffs a lower bar(Bidwell et al. 2022a). Some of the complexity in comparing the Bane Act to the state causes of action in our treatment states arises because there is no agreed upon test for determining exactly how to meet the statutory requirements laid out in the Bane Act. One line of state cases maintains that plaintiffs must "show threat, intimidation, or coercion other than that inherent in the alleged violation of constitutional or statutory rights" while another line rejects that standard and instead requires specific intent to violate a particular right of the victim. The specific intent approach requires that the right allegedly violated be "clearly delineated" and that the officer intended to deprive the victim of that right (Wolf 2020). The clear delineation requirement seems quite similar to the clearly established standard in §1983, though some commentators note that it can be applied at a level of generality which would be unacceptably broad under §1983 (Wolf 2020). The state supreme court has not cleared up which of these lines of case is correct, but it has noted that in the excessive force context, plaintiffs sometimes run into difficulties when trying to use the Bane Act.[25] Some limited action in favor of plaintiffs was taken in 2021 when the legislature amended the Bane Act to weaken immunities for certain law enforcement officers in the state (Gildin 2024).

[24] Two other recent reforms were more limited in scope than those described here. In 2021, the Massachusetts legislature passed a police reform bill which created a new mechanism for decertifying police officers and removed qualified immunity as a defense for officers who were decertified as a result of their misconduct (Gildin 2024). Because decertifications are rare, this reform will not have broad applicability across the policing profession.

The state of New York has not adopted qualified immunity reform, but New York City has repealed qualified immunity for police officers sued for unreasonable searches and seizures. Unlike the reforms in New Mexico, Colorado, and Nevada, this repeal of qualified immunity does not apply to suits arising under other parts of the state constitution (Sibilla 2021). While this municipal level repeal of qualified immunity is significant, the majority of the state is unaffected.

[25] *B.B. v. Cty. of L.A.*, 10 Cal. 5th 1, 32 (2020)

Connecticut

Around the same time Colorado and New Mexico passed their qualified immunity reforms, Connecticut enacted its own version. Connecticut law allows individuals to sue police officers for violations of their state constitutional rights, but immunizes the officers from liability when they have "an objectively good faith belief" that at the time they acted their conduct did not violate the rights of the victim (Gildin 2024). Commentators note that this carveout threatens to become an exception that swallows the rule and weakens the effectiveness of this reform (Sibilla 2020). However, at face value the good faith defense seems similar to the federal qualified immunity standard set up before *Harlow* and as such would be substantially different from the current form of the doctrine.[26]

Maryland

In Maryland, a complex system of immunities sometimes protects, and other times exposes the government and its officials to liability for their misconduct.[27] This system has been described as "relatively plaintiff friendly" and defies simple comparison to our treatment states of Colorado, New Mexico, and Nevada (Bidwell et al. 2022b).

Michigan

In a 2022 case *Bauserman v. Unemployment Insurance Agency*[28], the Michigan Supreme Court held that individuals can sue the state for damages under the state constitution when their rights are violated by state actors. Governmental immunity is not applicable to these claims (Bidwell 2023). While this judicial reform seems similar to Nevada's reform, it is not entirely clear that *Bauserman* is as powerful a rejection of government immunity in the law enforcement context. To start, the case does not deal with police misconduct, it addresses a due process violation committed by a state insurance agency. Furthermore, the court in *Bauserman* denied that their holding applied to individual government actors or municipal governments.[29] More fundamentally, Michigan law only allows suit for damages under the state constitution when no adequate alternative remedy exists for an injury. Case law indicates that the availability of §1983 as a means for remedying rights violations constitutes an adequate remedy, at least in some cases.[30] This means many police misconduct claims will be funneled to §1983, where they will face the hurdle of qualified immunity, and be barred from being brought under the state constitution. Other police misconduct claims might be channeled to the Michigan Governmental Tort Liability Act, but will face immunity difficulties there as well (Bidwell et al. 2022c).

Montana

As discussed in the main text, the Montana Supreme Court created a state work-around to federal qualified immunity in 2002 when it decided *Dorwart v. Caraway*[31]. Since Montana was under

[26] For a description of the qualified immunity standard pre-*Harlow,* see *Peirson v. Ray,* 386 U.S. 547 (1967)
[27] *Clea v. Baltimore*, 312 Md. 662 (1988); *Houghton v. Forrest*, 412 Md. 578 (2010)
[28] 509 Mich. 673 (2022)
[29] This is discussed in footnote 13 of the opinion.
[30] *Searcy v. Cty. of Wayne*, Nos. 365111, 365114, 365118, 2024 Mich. App. LEXIS 4256 (Ct. App. May 30, 2024)
[31] 58 P.3d 128 (2002)

the same liability regime during the entirety of our sample period its legal status does not impact our estimates; however, unlike the never treated control states, this state should be thought of as an always treated control state.