Public Benefit
Private Burden

A National Study of Compensation for Property Damage Caused by Law Enforcement

By David Warren, Ph.D.,
Jason Tiezzi, and
Mindy Menjou

September 2026

Institute for Justice

Public Benefit
Private Burden

A National Study of Compensation for Property Damage Caused by Law Enforcement

Institute for Justice

By David Warren, Ph.D., Jason Tiezzi, and Mindy Menjou

September 2026

Table of Contents

Executive Summary

In the course of their duties, law enforcement officers sometimes damage property belonging to innocent individuals. The damage could be as minor as a fence broken during a foot chase or as extensive as the destruction of a home where a suspect has barricaded himself. In either case, after the damage is done, the question becomes who pays the bill: the unlucky property owner or the government?

This is the first national study to quantify how often property owners seek compensation from local governments for damage caused by law enforcement—and how often governments honor those requests.

All told, we identified more than 2,700 claims for law enforcement property damage across 222 local jurisdictions from 2015 through 2023. Over the same nine-year period, that extrapolates to about 12,000 claims across the 1,027 local jurisdictions home to the nation's largest law enforcement agencies with SWAT teams, the population from which the 222 were drawn.

Most claims involved residential damage from routine law enforcement activities, though 12% involved tactical raids. Vehicle damage was common. Among all claims where the outcome could be determined, 60% were denied, and payment rates varied widely across jurisdictions.

Other key findings include:

- Unexpected property damage can create hardship for ordinary people. The median claim amount is $1,260, which is more than most Americans have available to cover an emergency expense. And some claims are far costlier: About 1 in 10 claims sought $10,000 or more.
- Unlucky bystanders are often left holding the bag. More than 90% of claimants were not the target of the law enforcement action that led to the damage, and more than 70% of those had no relationship to the target. Still, local governments denied 41% of these owners' claims.
- Compensating worthy claimants would pose little burden to local governments. Not only do some governments already pay most claims, but paying all claims in full—regardless of those claims' merits—would cost municipalities an average of only about $6,500 per year, which represents a negligible fraction of their annual spending. And although large-dollar claims do occasionally happen, individual jurisdictions generally see very few, if any, and even those represent tiny fractions of municipal spending.
- Compensating owners is unlikely to restrain law enforcement. In hundreds of claims, owners said they had been told by officers how to seek compensation, or even that the municipality would cover the damage, indicating both that these officers understood the costs from their actions could fall on the municipality *and* that they took action anyway.

The Fifth Amendment's takings clause requires the government to pay owners "just compensation" when the government takes their property for a public use, such as when the government uses eminent domain to take a home for a road or a school. An animating principle of the clause is protecting individual property owners from bearing burdens that should be borne by the wider public. Unfortunately, this unfairness is exactly what happens when governments refuse to compensate innocent victims of property damage by law enforcement: Private citizens are left to bear public burdens alone.

The U.S. Supreme Court will soon have an opportunity to rectify this situation. The Institute for Justice has asked it to hear the cases of two innocent owners whose properties were destroyed by SWAT teams trying to apprehend criminal strangers. Whether the Court takes up these cases or future ones, our results suggest that owners like these can be compensated for such takings without ill effects.

Introduction

Carlos Pena was living the American Dream. He owned a successful print shop in the North Hollywood neighborhood of Los Angeles that he hoped to one day pass on to his son.

Then, one summer day in 2022, Carlos' dream was dashed.

Carlos was working in his shop when he heard a commotion outside. He opened the door to investigate and saw a man barreling toward him, U.S. Marshals in hot pursuit. The fugitive grabbed Carlos, forced him out of the shop, and barricaded himself inside. The agents then ordered Carlos to stay back as they surrounded the shop.

Officers from Los Angeles' Special Weapons and Tactics team soon arrived. The SWAT officers fired dozens of tear-gas canisters into the shop, breaking windows and causing extensive damage to the structure and everything inside. All of Carlos' printing equipment, as well as his inventory, was destroyed. Finally, after a prolonged standoff, the officers entered the shop only to find that the fugitive had somehow escaped.

All told, the SWAT officers caused more than $60,000 worth of damage. But when Carlos sought compensation from the city of Los Angeles, it refused to consider his claim.[1]

Texas homeowner Vicki Baker faced a similar ordeal in the city of McKinney, Texas. In 2020, police destroyed her home while trying to apprehend a fugitive who had taken a teenage girl hostage. Like Los Angeles, McKinney refused to pay for the damage law enforcement had caused in the course of its duties.[2]

In both cases, cities chose to leave unlucky innocent property owners holding the bag for damage deliberately caused by law enforcement carrying out its mission to protect the public. And insurance was not an option for Carlos or Vicki, as their policies, like most, exclude damage caused by government actors.[3]

Until now, no national study has examined how often people like Carlos and Vicki seek compensation for damage caused by law enforcement, nor how often local governments honor their claims. The closest analog is "Damaged for Good," a recent series from the *Austin American-Statesman*, which looked at claims in the Texas state capital and found 135 claims for building damage caused by police over a six-year period, all of which the city refused to pay.[4]

This study takes a similar approach, but with a nationwide sample of local jurisdictions.[5] (We also sought records from state and federal law enforcement agencies, but many have not responded. This study therefore focuses only on municipalities.) In all, we identified more than 2,700 claims across 222 jurisdictions from 2015 through 2023. Extrapolating to the 1,027 local jurisdictions in the United States with the largest law enforcement agencies that maintain SWAT teams—the population from which the 222 were drawn—we estimate there were about 12,000 claims over that same nine-year period.

While many of these claims, like Carlos' and Vicki's, involved SWAT teams and thousands of dollars in damage, the typical claim was much smaller—and yet local governments often refused to pay.

The median claim amount was $1,260, a tiny amount in the scheme of municipal budgets but more than most Americans have on hand to cover an emergency expense.[6] Claims most commonly pertained to residential damage that did not involve a raid or pursuit. Typical examples include police breaking down a door or fence to search for a suspect or respond to an emergency. Vehicle claims were also frequent, with a common example being police crashing into another motorist while responding to a call.

Among all claims where the government's decision to compensate could be determined, 60% were denied. We also found that most claimants had nothing to do with the law enforcement action that led to their property's damage—and still their claims were often denied. Where claimants were neither the target of the action nor had any relationship with the target, local governments still denied 41% of claims. And whether a claimant received compensation varied widely across jurisdictions. Some jurisdictions paid almost every claim, while others paid almost none.

In 2024, the U.S. Supreme Court had an opportunity to ensure worthy claimants are made whole when it considered whether to hear a constitutional case brought on Vicki's behalf by the Institute for Justice. Vicki argued that by denying her claim, the city of McKinney violated the Fifth Amendment's takings clause, which allows the government to take private property for public use—if the government pays owners "just compensation." Much like people are due just compensation if the government decides it must take their home to build a road or school, Vicki argued innocent owners like her are due compensation when law enforcement deliberately damages or destroys their property—in effect, "takes" it—for the public good, such as apprehending a fugitive.[7]

Though the Court declined to hear Vicki's case, Justices Sotomayor and Gorsuch acknowledged that her home had been destroyed for a "public benefit" and noted that she "undoubtedly would be entitled to compensation" if McKinney had "razed [her] home to build a public park."[8] Quoting an earlier Supreme Court opinion, they also pointed out that the "Takings Clause was 'designed to bar Government from forcing some people alone to bear public burdens which, in all fairness and justice, should be borne by the public as a whole.'"[9]

Our analysis suggests municipalities could bear such public burdens with little trouble. In a given year, the average jurisdiction faced just one or two property damage claims. On average, if the municipalities in our study had to pay every property damage claim in our dataset, the average jurisdiction would have been on the hook for only about $6,500 per year. And in reality, they likely would not have to pay every claim in full, as some would not count as takings and cities sometimes settle claims for lesser amounts. Moreover, our claims data suggest officers often understood the costs of their actions could fall on the municipality, not the property owner. We found hundreds of examples of officers telling

owners how to file claims after damage had occurred and in some cases even suggesting the municipality would pay for the damage.

Justices Sotomayor and Gorsuch observed that the Court might eventually have to take up a case such as Vicki's, and it will soon have another chance.[10] IJ has asked it to hear both Carlos' case and that of Amy Hadley, whose South Bend, Indiana, home was destroyed by police during a 2022 "wrong house" raid.[11] Whether in their cases or in future litigation, our analyses suggest there is good reason for the Court to prevent public benefits from becoming private burdens.

Methods

The objective of this study was to obtain and analyze property damage claims submitted to local governments for damage caused by law enforcement. We were concerned only with claims that resulted from law enforcement activities. Most notably, this means we were not interested in automobile accidents where police were driving in traffic like a regular motorist (e.g., returning from a call back to the police station). We attempted to obtain claims covering the period 2015 through 2023 with the goal of answering the following questions:

1. How many claims for property damage were filed?
2. What was the value of a typical claim?
3. What types of incidents led to claims?
4. What types of individuals or entities filed claims?
5. What were the outcomes of claims?
6. How did claim payment rates vary by jurisdiction?

With these questions in mind, we identified the largest 20% of law enforcement agencies with SWAT teams in the United States, a total of 1,027 agencies.[12] We then randomly selected 467 of these agencies and sent public records requests to their parent jurisdictions for any claims filed against the agency alleging property damage caused during law enforcement actions.[13] We also requested all records relating to such claims, including the government's decision on, or "determination" of, whether to pay a claim.

We received responses from 308 agencies. Of these, 86 provided only spreadsheets listing claims. Because most spreadsheet responses contained limited or incomplete information, we did not use them for this analysis.[14] Another 178 jurisdictions provided documentation for each claim, including completed claim forms, receipts, correspondence, and, often, police reports and claim determination documents. For these jurisdictions, we separated each claim into its own PDF file, resulting in roughly 4,500 individual claim records totaling approximately 80,000 pages. The remaining 44 jurisdictions indicated they had no responsive records; these are included among the 222 jurisdictions.[15]

Given the time it would take IJ researchers to review 80,000 pages, we instead used artificial intelligence to identify relevant claims and code them across additional variables. This section describes our study variables and how we used human-coded records to train the AI coder and ensure its reliability.

VARIABLES OF STUDY

For each claim record, we first coded a threshold question of whether the record was relevant for our study. To be considered relevant, a claim record needed to contain evidence that the claimant was seeking reimbursement for property damage caused by law enforcement during a law enforcement

activity. Claim records were most commonly considered irrelevant because they (1) included only bodily injury claims; (2) involved routine motor vehicle accidents where the law enforcement officer was in traffic like any other driver; or (3) contained little to no information about the underlying incident that led to the damage.

If a record was relevant, we then coded additional variables that described various attributes of the claim, including the type of property damage claimed, when the claim was filed, and whether the claim was ultimately paid or denied. A list of the variables included in our study is provided in Table 1, while full definitions for each variable are available in our codebook in Appendix C.[16] The codebook also lists the kinds of claims we excluded as irrelevant.

Table 1: Study variables

	Field	Description	Type	Response Options
Relevance	Relevance	Whether the claim record was relevant for our study (i.e., whether it involved property damage caused by law enforcement during a law enforcement activity)	Binary (Y/N)	--
Claim Basics	Claim amount	The value of the property damage claimed	Numerical	--
	Deductible only	Whether the claimant sought reimbursement for an insurance deductible only	Binary (Y/N)	--
	Incident date	The date the incident that led to the claim occurred	Date	--
	Claim date	The date the claim was submitted	Date	--
	Type of damage	The type(s) of property damage that were included in the claim	Categorical (select all that apply)	A - Residential property B - Commercial property C - Agricultural property D - Vehicle E - Personal property
Claimant Information	Claimant type	The type of person or entity submitting the claim	Categorical	A - Individual for themselves B - Lawyer C - Individual for someone else D/E - Insurance company* F - Business
	Claimant was target	Whether the claimant was the reason for the law enforcement action that led to the damage	Categorical	A - Yes B - No C - Unknown
	Relationship to target	If the claimant was not the target, the nature of the relationship between the claimant and the person who was the reason for the law enforcement action	Categorical	A - No relationship B - Landlord C - Parent D - Other family E - Friend or romantic partner F - Other relationship G - Unknown

	Field	Description	Type	Response Options
Outcome Information	Claim outcome	Whether the claim was paid or denied by the local government	Categorical	A - Denied B - Paid C - Unknown
	Paid amount	The amount paid to the claimant by the government, if applicable	Numerical	--
	Reason for denial	The reason(s) that a claim was denied, if applicable	Categorical (select all that apply)	See codebook in Appendix C
	Determination date	The date of the government's determination decision	Date	--
Incident Circumstances	Pursuit	The damage was caused by a law enforcement pursuit	Binary (Y/N)	--
	Tactical raid	The damage was caused by a SWAT/tactical raid	Binary (Y/N)	--
	Spike strips	The incident involved spike strips set out for a different driver	Binary (Y/N)	--

* Although there were initially two categories for insurance companies, we consolidated them into one for the majority of our analyses. Response option D was an insurance company filing on behalf of an insured customer; E was an insurance company filing on behalf of itself.

TRAINING THE AI CODER

To train our AI coder, we created training and validation datasets with 86 records each. IJ researchers manually coded these datasets across all study variables. Once the datasets were human coded and quality checked, we used them to improve the AI coder's performance. Further details about our processes for both ensuring the quality of our human coding and training the AI model can be found in Appendix A.

EVALUATING THE AI CODER'S PERFORMANCE

Once the training process was complete, we evaluated the AI coder's performance on a test dataset composed of 200 human-coded records unseen during the training process. The AI coder performed well, achieving an overall accuracy rate across all questions of 94.7%. Its performance was in line with that reported by several other legal and public policy studies that use AI coding of a dataset.[17] Table 2 lists the performance for each field in our study using the most relevant performance metric. For numerical, date, and categorical (i.e., multiple choice) fields, we used accuracy as the primary performance metric. For binary (yes/no) fields, we used a statistic called the F1 score. F1 scores range from 0 to 1, with 1 being perfect.[18] For more detailed performance information, see Appendix B.

FINALIZING THE DATASET

Once our AI coder was finalized, we used it to code the approximately 4,000 uncoded claims and combined that coding with the human-coded claims from the training, validation, and test datasets. We then performed minor clean-up and quality control, including manual review of outliers and internal inconsistencies. We also replaced 17% of AI-coded claim outcomes with verified outcome information that some jurisdictions provided in spreadsheet form. This quality control is described in more detail in Appendix A. Based on this quality control, our reported performance statistics likely understate the reliability of our final dataset, especially for fields involving dollar values and claim outcomes.

Our final dataset contains 2,752 relevant claims and over 50,000 individual datapoints. A copy of this dataset can be found on IJ's website at https://ij.org/report/public-benefit-private-burden/.

Table 2: Performance statistics

Field Type	Field	Primary Performance Statistic	Performance
Numerical / Date	Claim amount[a]	Accuracy	91.2%
	Incident date	Accuracy	97.4%
	Claim date	Accuracy	93.9%
	Paid amount[a]	Accuracy	93.0%
	Determination date	Accuracy	92.1%
Categorical	Type of damage	Accuracy	94.7%
	Claimant type	Accuracy	89.5%
	Claimant was target	Accuracy	91.2%
	Relationship to target	Accuracy	89.5%
	Claim outcome[a]	Accuracy	95.6%
	Reason for denial	Accuracy	90.4%
Binary (Y/N)	Relevance	F1 Score	0.938
	Deductible only	F1 Score	N/A[b]
	Pursuit	F1 Score	0.927
	Tactical raid	F1 Score	0.971
	Spike strips	F1 Score	1.000

[a] We made additional manual corrections for these fields. See Appendix A for details.

[b] There were no-deductible only claims in the test dataset; we were therefore unable to calculate an F1 score, although accuracy was perfect at 100%.

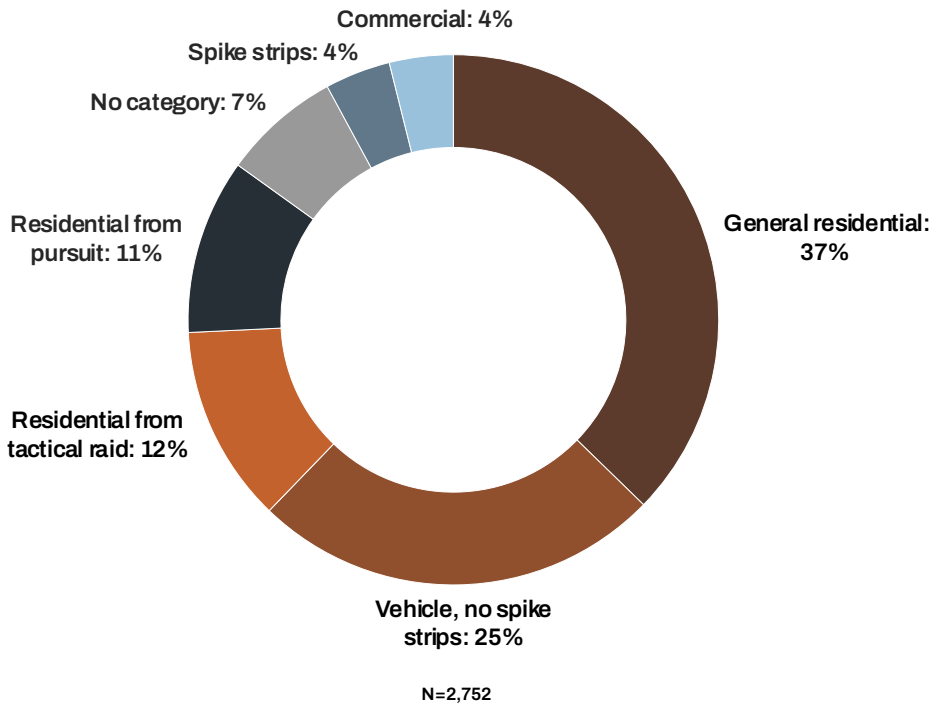
Results

THE SCOPE OF PROPERTY DAMAGE CLAIMS

From 2015 through 2023, there were 2,752 claims for property damage caused by law enforcement across the 222 responding jurisdictions. That figure includes about 400 claims where records indicated raids by SWAT or other tactical teams. Based on these figures, we infer that over the same nine-year period, 12,000 claims (±3,000) were filed with the 1,027 local jurisdictions in the United States with the largest law enforcement agencies that maintain SWAT teams. This includes an estimated 1,600 (±300) claims where damage was caused by tactical teams.[19] These figures likely understate the true extent of property damage, as some owners may forgo filing a claim because they do not know the process or for other reasons.

We classified claims into six broad categories based on the type of property damaged and the circumstances of the incident.[20] As shown in Figure 1, most claims were for damage to residential property, most commonly incidents where records did not indicate a pursuit or a tactical raid (37% of all claims). A typical example of such general residential property damage was police forcing entry into a residence for a search and damaging the front door. Residential damage claims that indicated a tactical raid made up another 12% of claims, while residential damage resulting from pursuits, such as a fence broken during a foot chase, made up 11%.

Figure 1: Property damage claims by type



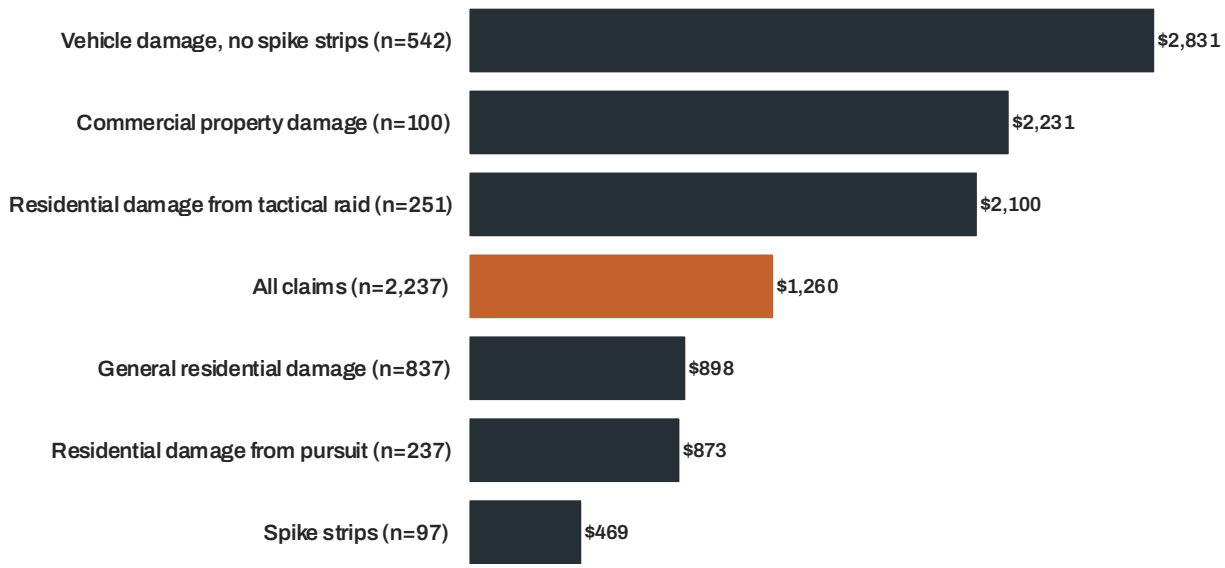
Vehicle damage was also common, comprising 29% of claims. Four percent were for vehicles damaged by spike strips, which law enforcement sometimes deploys in vehicle pursuits. The other 25% of vehicle damage claims almost always involved motor vehicle accidents that occurred when officers were responding to an incident.[21]

Relatively few claims were for damage to commercial property (4%), such as damage to a business during the arrest of a patron. We left uncategorized the roughly 7% of claims that did not fall neatly into a category.

THE SIZE OF PROPERTY DAMAGE CLAIMS

The median value of a property damage claim in our data was $1,260. As shown in Figure 2, this value varied by the broad claim type, although all were under $3,000. Vehicle damage claims (excepting those from spike strips) were the most expensive, with a median value of $2,831. Residential damage claims were typically smaller, reflecting the fact that such claims usually involved relatively minor damage. Medians were under $900, with the notable exception of claims involving tactical raids.[22] As might be expected, the median tactical raid claim was much larger at $2,100.[23]

Figure 2: Median amount requested by claim type



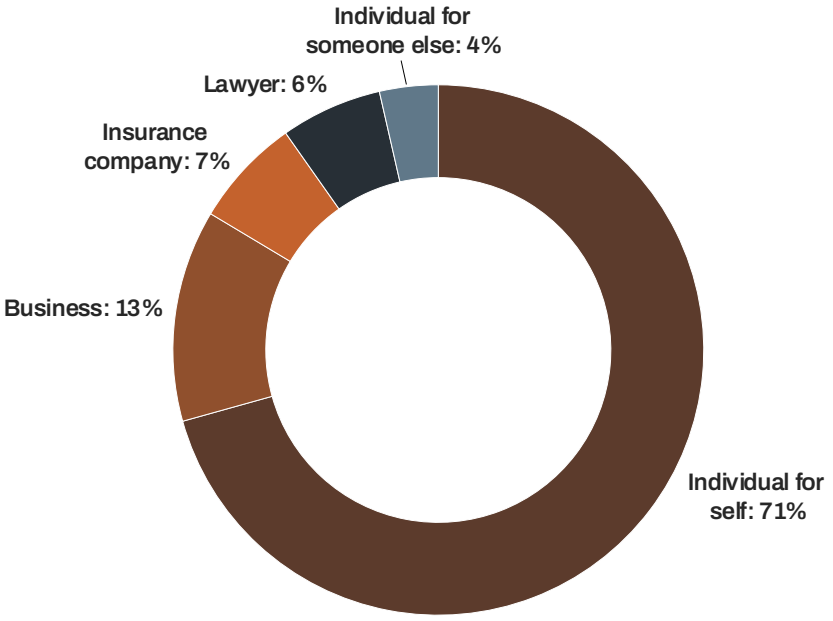
Though typical claims are relatively small, some—like Carlos Pena's and Vicki Baker's— involve tens of thousands of dollars in damage. About 1 in 10 claims were for $10,000 or more, and about a quarter of those larger claims indicated the damage resulted from tactical raids.

For example, the owner of an appliance shop in Tulsa, Oklahoma, filed a claim for $35,485.69 after a standoff between SWAT officers and a gunman left his shop in shambles. And in Los Angeles County, a family sought $37,260 after officers used their property to execute a SWAT raid on the house next door, causing extensive damage in the process.[24] Like Carlos' and Vicki's, both these claims were denied.

WHO FILES PROPERTY DAMAGE CLAIMS

Most claimants (71%) were individuals filing on their own behalf. Other claimant types, which included businesses (13%), insurance companies (7%), and lawyers (6%), were less common.[25] (See Figure 3.)

Figure 3: Most claimants were individuals filing on their own behalf

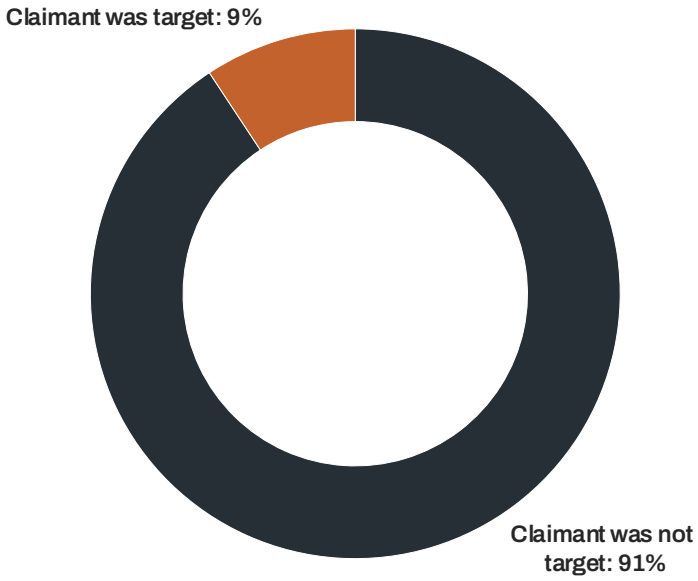


Note: Percentages do not sum to 100% due to rounding. (N=2,748)

In addition, records indicated that claimants were rarely the target of the law enforcement action that led to the damage—and most often, they had no relationship with the target. Among records with sufficient information, 91% of claimants were not the reason for the law enforcement action (see Figure 4).[26]

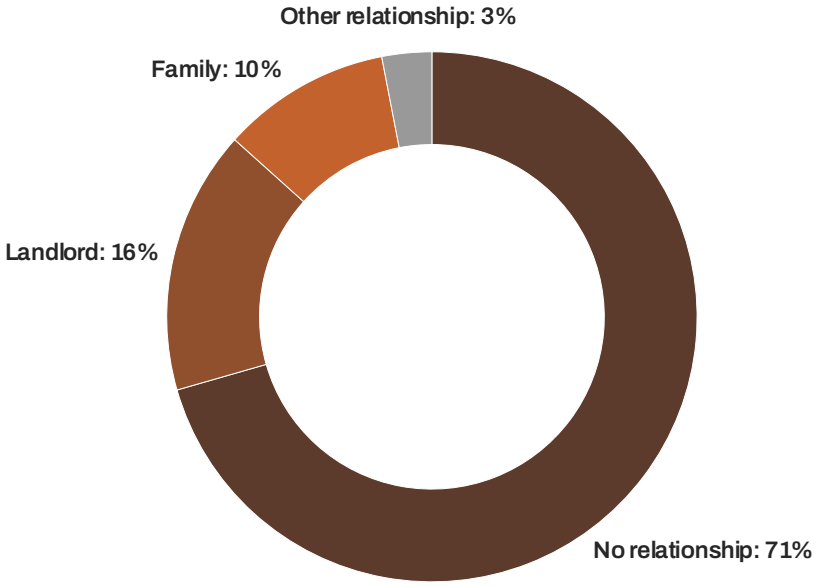
Similarly, among records where the claimant was not the reason for the law enforcement action and there was sufficient information, 71% of claimants had no relationship with the law enforcement target.[27] When there was a relationship, the claimant was most often the target's landlord (16%) or family member (10%). (See Figure 5.)

Figure 4: Claimants were rarely the target of the law enforcement action that led to the property damage



Note: Data include all claims where the target of the law enforcement action could be determined. (N=2,197)

Figure 5: Claimants typically had no relationship with the target of the law enforcement action

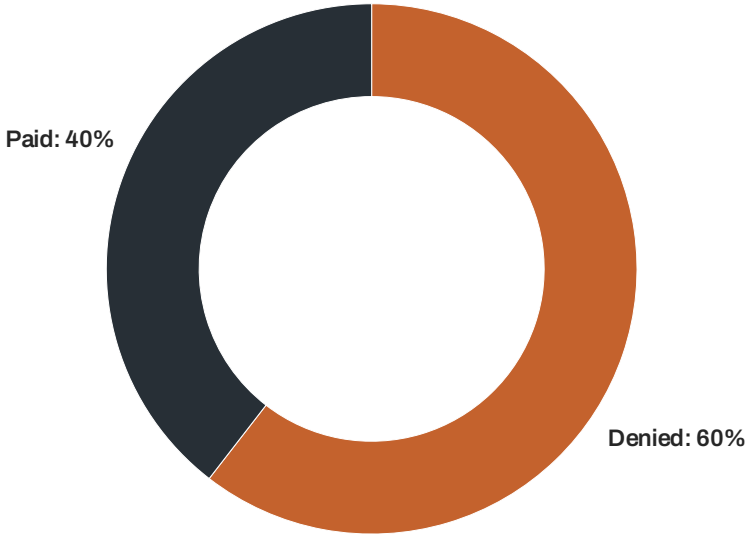


Note: Data include all claims where the claimant was not the target of the law enforcement action and the relationship with the target could be determined. (N=1,813)

THE OUTCOME OF PROPERTY DAMAGE CLAIMS

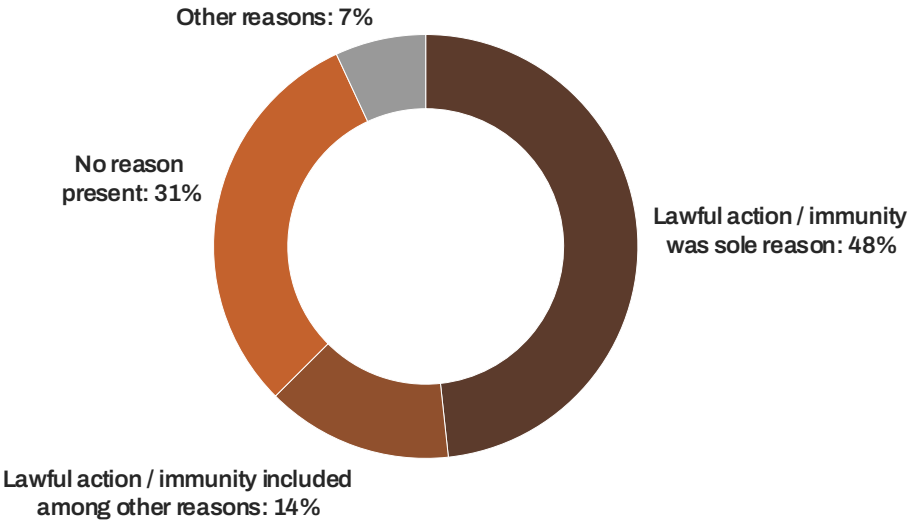
Among claims where the outcome was recorded, 60% were denied, while 40% were paid (see Figure 6).[28] In nearly a third of denials (31%), the government either gave no reason for denying claims or the reason could not be determined from the records provided. In another 48%, the government's sole reason for denial was that the law enforcement officers' actions were legal and within the scope of their employment or that the government was immune. (See Figure 7.) Often, this was boilerplate language that did not wrestle with the facts of a situation. Regardless, the frequency of such denials suggests governments often conflated the question of whether an officer's action was lawful with the question of whether compensation was owed.[29]

Figure 6: Most property damage claims were denied



Note: Data exclude claims with no reported outcome. (N=2,303)

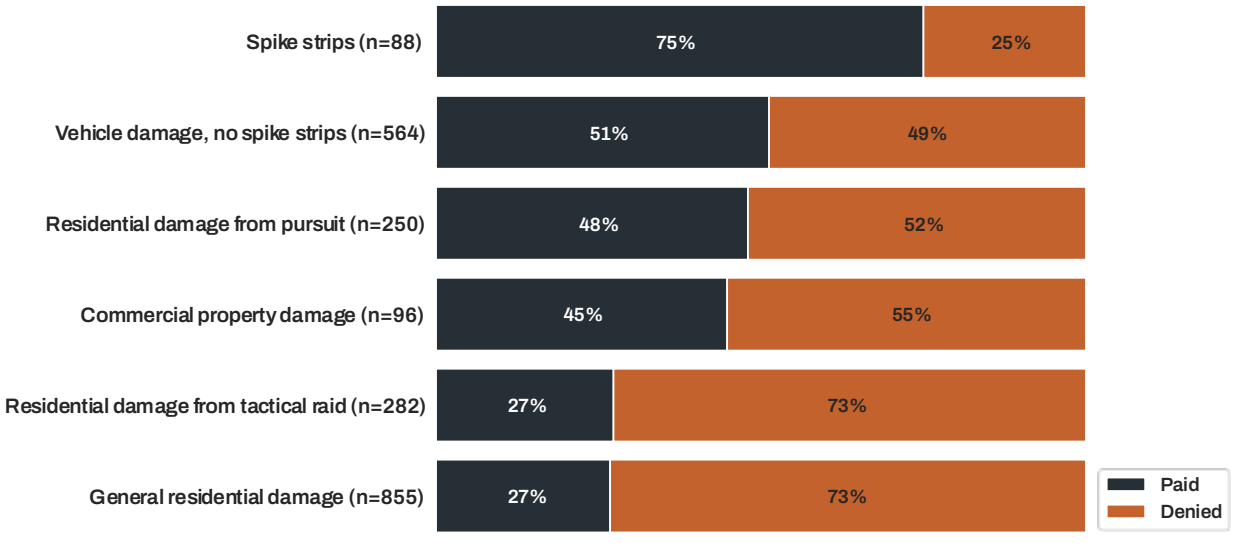
Figure 7: When claims were denied, jurisdictions typically claimed that officers' actions were lawful or that the government was immune



Note: Data exclude roughly 300 claims where outcome information was provided separately without denial reasons. (N=1,071)

While claims were denied more often than they were paid, the payment rate varied based on the claim type. Those involving vehicle damage were more likely to be paid than not, especially when spike strips were involved. Meanwhile, claims for residential damage resulting from tactical raids and general residential damage were nearly three times more likely to be denied than paid. (See Figure 8.)

Figure 8: Outcomes by claim type



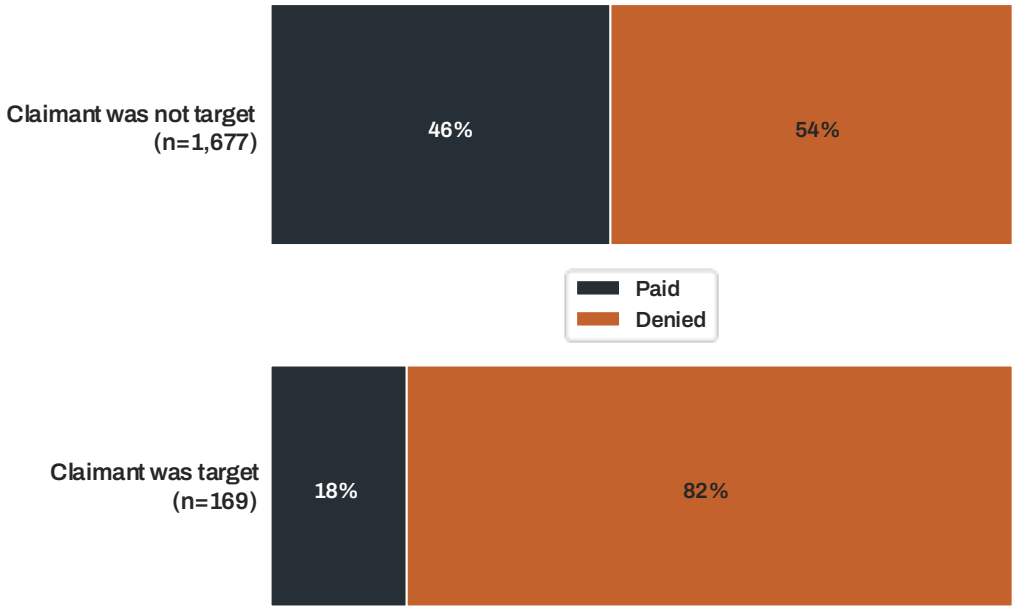
Note: Data exclude claims with no reported outcome.

Perhaps not surprisingly, municipalities more frequently compensated claimants who had no connection to the law enforcement action—those who were not a target or who had no relationship to the target—but many such claims were denied.

Claims were far less likely to be paid when those seeking payment were the target of the law enforcement action. Just 18% of such claims were paid. By contrast, municipalities paid 46% of claims made by those who were not a target. Still, 54% of claims made by non-targets went uncompensated. (See Figure 9.)

Figure 9: Outcomes by whether the claimant was the law enforcement target

Even when the claimant was not the target, 54% of claims were denied

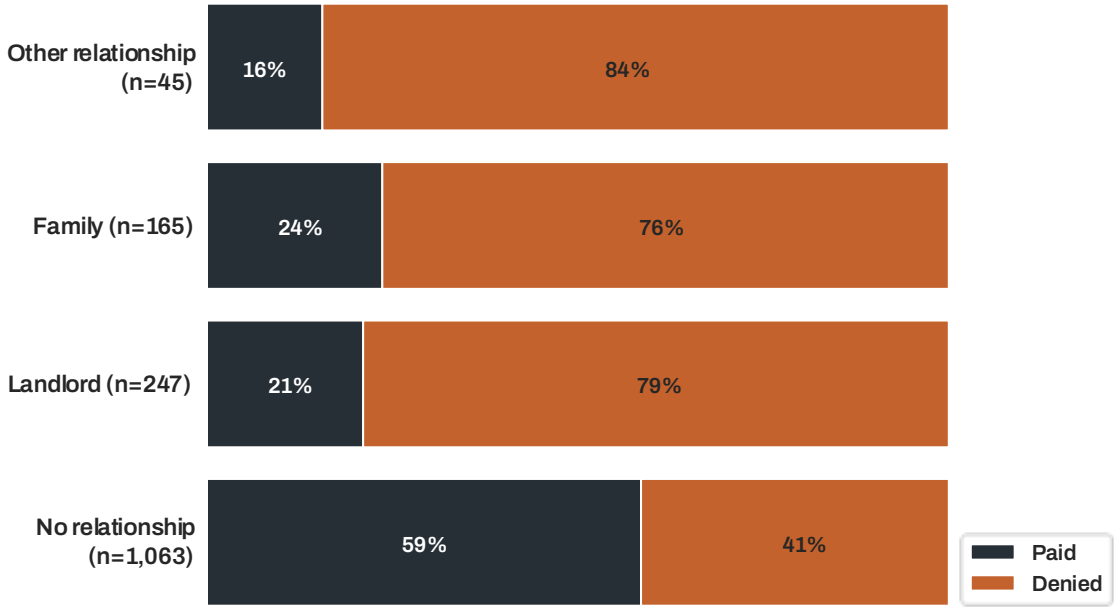


Note: Data exclude claims with no reported outcome.

Similarly, claims were far less likely to be paid if the claimant bore some relationship to the target, such as a family member or landlord. Municipalities paid less than a quarter of such claims. They compensated unrelated claimants far more often, 59% of the time, but still left 41% of those without a relationship to the target unpaid. (See Figure 10.)

Figure 10: Outcomes by the claimant's relationship to the law enforcement target

Even when the claimant had no relationship with the target, 41% of claims were denied



Note: Data exclude claims with no reported outcome.

One factor that does not appear strongly related to payment is the amount of the claim. While lower-value claims were slightly more likely to be paid than higher-value ones, the relationship is weak. For example, while the median claim amount for denied claims exceeds that for paid claims ($1,301 versus $1,128), the difference is relatively small. In addition, when we performed a logistic regression analysis controlling for factors such as the claimant's relationship with the law enforcement target, as well as characteristics of incidents, claims, and jurisdictions, we found only a small, albeit statistically significant, relationship between claim amount and payment decision: Doubling the claim amount from $1,000 to $2,000 is associated with a 2.5 percentage point increase in the probability of denial.[30]

PAYMENT RATES BY JURISDICTION

Jurisdictions varied substantially in whether they paid or denied property damage claims. Among jurisdictions with more than 20 decided claims, Salinas, California, and Arvada, Colorado, represent opposite ends of the spectrum. Salinas paid only three of 39 decided claims, a payment rate of 8%. Arvada, meanwhile, had a payment rate of 96%. It paid 26 of 27 decided claims. A full third of jurisdictions with at least 10 decided claims had payment rates of under 30%, while 14% had payment rates of over 70%.

This variation holds even when we compare similar types of claims. For example, looking at residential damage resulting from tactical raids where the claimant was not the target, we see jurisdiction payment rates ranging from 0% in El Paso County, Colorado, and the District of Columbia to 80% in Gardena, California (see Figure 11).[31]

Figure 11: Payment rates for residential damage claims from tactical raids when the claimant was not the target

Note: Data are from jurisdictions with at least six decided claims for residential damage from tactical raids where the claimant was not the target of the raid.

We see similar variation in payment rates for residential damage resulting from pursuits where the claimant was not the target. Two jurisdictions (Milwaukee, Wisconsin, and the District of Columbia) did not pay a single such claim, while four paid over 80% (see Figure 12).[32]

Figure 12: Payment rates for residential damage claims from pursuits when the claimant was not the target

Note: Data are from jurisdictions with at least six decided claims for residential damage from pursuits where the claimant was not the target of the chase.

THE BURDEN ON MUNICIPALITIES

Though our data indicate that thousands of property owners have sought compensation for property damage caused by law enforcement, we find the financial burden of paying such claims for any given municipality in any given year would be light.

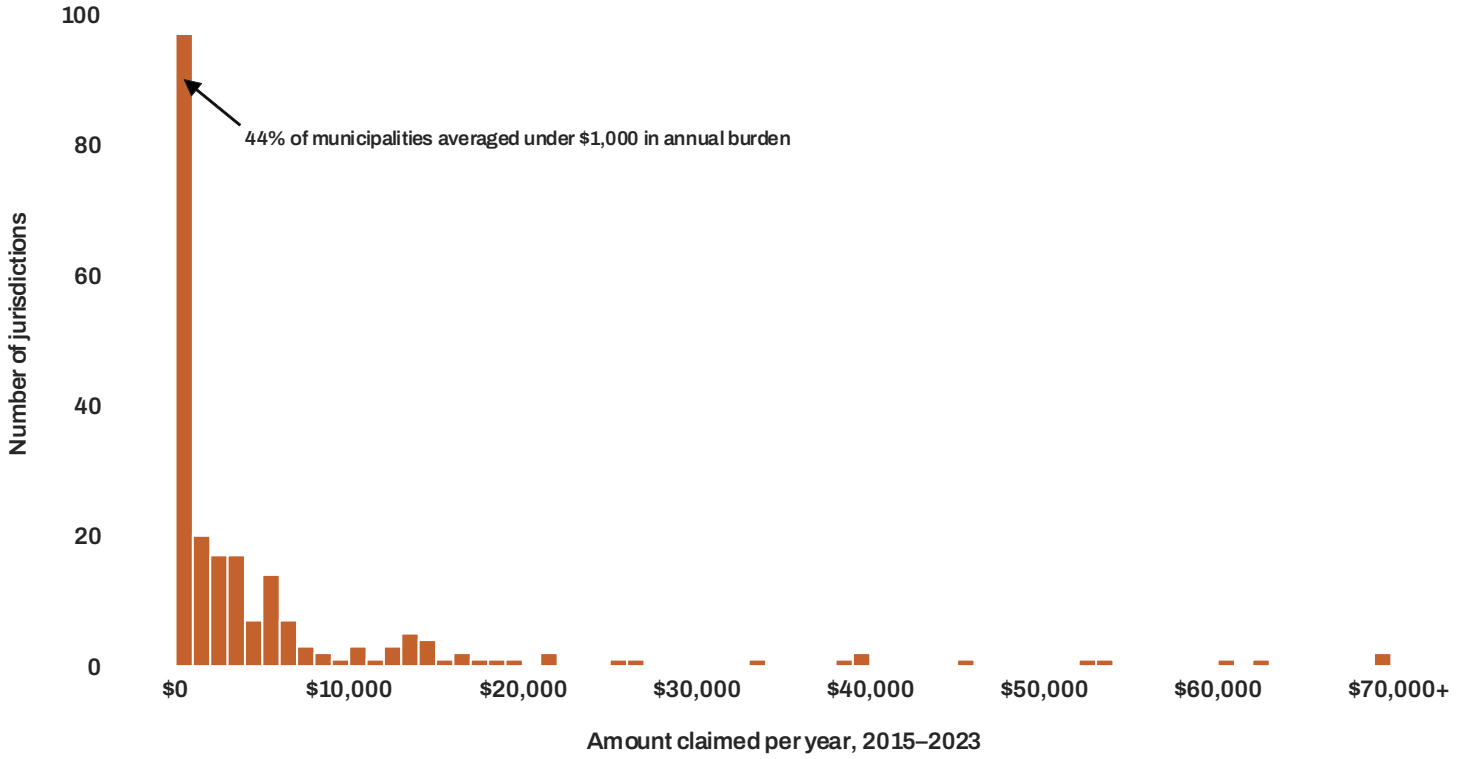
For starters, municipalities typically receive very few property damage claims annually—on average, one or two per year.[33] In our nine-year dataset, 44 jurisdictions received none. And even those with relatively high volume saw only a couple dozen per year. The District of Columbia, for example, led the pack with an annual average of 25.

The total financial burden of these claims is also low. If every property damage claim in our dataset from 2015 through 2023 had been paid in full, the total amount paid across the 222 jurisdictions would have been $12.94 million over nine years, or an average of $6,476 per jurisdiction per year.[34] The median burden is smaller, just $1,675 per year, as the average is skewed by a few jurisdictions with higher-value claims. (See Figure 13.) For example, Federal Way, Washington, had one of the highest average claim amounts at about $63,000 per year, yet 85% of its burden came from a single tactical raid that badly

damaged a house.[35] The claimant sought $482,000 but later settled for $60,000. Without that claim, the city's average would be $9,443 per year.

To be sure, not all claims in our dataset would be considered "takings" of private property for public use such that the Fifth Amendment would require just compensation. For property damage to be a taking, the damage caused by law enforcement would have to be intentional and the owner could not be at fault for triggering the government action. This is necessarily a case-specific analysis that we cannot undertake for each of the 2,700 claims in our dataset. For that reason, the burdens presented here represent an estimated upper bound.

Figure 13: Estimated annual municipal burden of fully compensating all property damage claims in our dataset



Discussion

Whether breaking down a door to search for a suspect, puncturing a person's tires with spike strips meant for someone else, or deploying tear-gas grenades in a home or business to flush out a fugitive, law enforcement sometimes damages innocent people's property while trying to protect the wider public. When the government takes legitimate actions in the public's interest and property winds up damaged, the question becomes: Who should be left with the bill?

As the Supreme Court has long recognized, one of the principles animating the Fifth Amendment's takings clause is protecting unlucky property owners from bearing burdens that should be borne by the public as a whole.[36] Our results illustrate the wisdom of this principle by showing how unexpected property damage can create genuine hardship for people who did nothing wrong. They also show that compensating them would not strain local governments' finances, nor would it deter law enforcement from acting in the public interest.

UNEXPECTED PROPERTY DAMAGE CAN CREATE GENUINE HARDSHIP FOR ORDINARY PEOPLE

At $1,260, the median claim amount may not seem like much, but it likely represents a substantial burden for many claimants. To put it into perspective, it is within the range of what a typical family of four spends on groceries in a month and almost equals the national median monthly rent with utilities.[37] Not only that, but only 47% of Americans report having $1,000 on hand to cover an emergency.[38] In short, for many, if not most, people, an unexpected $1,300 expense could be devastating. And, of course, half of claims are larger—some much larger.

Beyond financial hardship, sudden property damage can impose other burdens on innocent owners, throwing people's lives into disarray. For example, IJ client Amy Hadley and her teenage son were forced to sleep in her car for days before her house was safe to enter after police bombarded it with tear-gas grenades. Amy then had to spend days cleaning up the mess.[39] And four years after the SWAT raid that destroyed his shop, Carlos is still working out of his garage with just a single printer.[40] As a result, he has lost out on tens of thousands of dollars in revenue.[41]

Even something as simple as a broken door can cause upheaval, as the *Austin American-Statesman*'s analysis of law enforcement property damage in the Texas state capital illustrated. After police mistakenly broke down his apartment door, not only did Eniekenimi Seifere have to pay for the repairs, but he also stayed home from work until repairs could be made. Seifere, who works the late shift at an assisted living facility, did not want to leave his family alone at night without a door.[42] The *Statesman* reported other renters found themselves at risk of eviction if they could not cover the cost of fixing doors broken by police.[43] One woman, a domestic violence survivor, reportedly declined to call 911 on her abuser because she felt she could not risk having police break her door after they had already done so

three months before: "'I don't have money to pay my rent. I don't have money to pay my light bill. I damn sure ain't got money to pay to get another door fixed. I already had to pay to fix the last door they took down.'"[44]

UNLUCKY BYSTANDERS ARE OFTEN LEFT HOLDING THE BAG

The records we collected indicate most people who sought compensation were unlucky individuals who bore no responsibility for the law enforcement action that damaged their property—and yet many of them were left empty handed. By far the most common claimants were individuals filing on their own behalf. More than 90% of claimants were not the law enforcement target, and more than 70% of those had no relationship with the target.

Some of these claimants even expressed understanding for law enforcement. The *Statesman* reported such sentiments among property damage victims covered in its series.[45] And our own dataset includes similar instances of claimants saying they understood the damage to their property was unavoidable:

- A Davenport, Iowa, claimant wrote: "[A] stolen car was PIT [precision immobilization technique] maneuvered into my retaining wall. . . . I know police were doing their job, but my [damaged] retaining wall is the result of that." The claimant requested $400 for the damage.
- A property manager in Westminster, Colorado, requested compensation for an apartment door broken in a SWAT raid, writing: "I understand the need for what was done, but was hoping there could be some sort of reimbursement for this door." The claimant requested $1,100 for the damage.
- A driver in Madison, Wisconsin, seeking payment for two tires destroyed by spike strips intended for another, included in his claim a note of thanks to police "for keeping us safe": "I 100% support our men and women in blue." The claimant requested $335.49.

Notwithstanding these owners' innocence, understanding, and reasonable desire to be made whole, their claims were all denied—and they are not alone. More than half of claimants who were not targets had their claims denied. And while some of these may have arguably borne some responsibility by virtue of some relationship with the target, municipalities still denied 41% of claims where the victim was neither the target nor had any relationship with the target. In short, our data suggest many people are being left to foot the bill for damages caused by law enforcement activity that they had nothing to do with.

Underscoring the problem, by far the most common reason municipalities gave for denying claims was that law enforcement's actions were lawful or that the government was immune. But such justifications miss the point. In cases like those of Carlos Pena, Vicki Baker, Amy Hadley, and other

blameless bystanders, just compensation is owed precisely because officers took lawful action in the public interest. When they do, the Fifth Amendment is intended to prevent unlucky owners from holding the bag.

Such owners may be especially unlucky if they happen to live in jurisdictions that routinely deny claims, including those quite similar to claims paid elsewhere. For example, if you live in Sacramento and police damage your fence while apprehending a suspect, the city's claims administrators are likely to compensate you, as they did for two claims in our data. But if you live in Milwaukee, where the city denied two comparable claims, you would almost certainly be out of luck. Similarly, Westminster, Colorado, paid $220 after police broke a fountain while pursuing a suspect through the claimant's backyard, but the District of Columbia refused to pay $415 for flowerpots and a downspout destroyed during a backyard chase. As Figures 11 and 12 show, such variation is common, even when limited to cases where claimants were not targets of the action.

COMPENSATING WORTHY CLAIMANTS WOULD POSE LITTLE BURDEN TO LOCAL GOVERNMENTS

While many individuals would have a hard time absorbing an unexpected $1,300 expense, let alone one many times larger, such costs are negligible compared to the resources at the disposal of local governments. On average, the 222 local governments in our dataset faced only about $6,500 annually in property damage claims.

This estimate amounts to just 0.00052% of the average amount of direct general expenditures among those same local governments, essentially a rounding error.[46] This is consistent with other research finding that the costs local governments incur from all claims and lawsuits against them generally account for a very small share of their expenditures.[47]

Even very large claims are unlikely to make a dent in local government budgets. Such claims are quite rare and tend to be dispersed across jurisdictions and over time. Less than 3% (79 claims) were $20,000 or more, and those were spread across 43 jurisdictions and all nine years of our study. Similarly, only about half a percent were $100,000 or more (15 claims), and those were spread across 12 jurisdictions and eight years. In addition to being uncommon, they are by no means unaffordable. To take an extreme example, at roughly $143,000, the District of Columbia had the highest average annual claim amount in the study. Yet this represents only 0.0008% of its nearly $19 billion in annual expenditures.[48]

Indeed, our results suggest that claim size is, at most, a minor factor in local governments' decisions to pay or deny a claim, as doubling the claim amount from $1,000 to $2,000 is associated with only a slight increase in the probability of denial—just 2.5 percentage points. Our dataset, in fact, contains numerous examples of larger claims being paid and smaller ones being denied within the same jurisdiction. For example, even as Federal Way, Washington, paid $60,000 for the destruction of a house in a SWAT raid, it denied a $216 claim for damage from a spike strip meant for someone else, even

though a police officer instructed the owner to file a claim. Similarly, Springfield, Ohio, paid a $2,402 claim for damage incurred to a home when a SWAT team used it to attack a neighboring house but denied a $100 claim submitted by an innocent property owner whose yard was damaged during a pursuit.

On the other hand, some local governments appear to have no problem paying most, or even all, of the claims they receive. Among jurisdictions with at least 10 decided claims, 14% paid over 70% of claims. Those jurisdictions were led by Newburgh, New York, and Escambia County, Florida, which paid all of their claims (17 and 14, respectively), and Arvada, Colorado, and Livermore, California, which paid all but one of their claims (26/27 and 11/12).[49]

Further showing that local governments can compensate law enforcement property damage, since 1998, Minnesota law has required local governments to pay compensation to innocent third parties whose property is damaged by law enforcement actions involving warrant service or apprehension of suspects.[50] In 2026, the Legislature bolstered the law by making it easier for eligible claimants to seek just compensation, particularly when tear gas and other damages from tactical raids are involved.[51] Also in 2026, and apparently at least partly in response to the *Statesman*'s reporting, Austin, Texas, announced it would begin compensating blameless victims of law enforcement property damage.[52]

COMPENSATING OWNERS WOULD NOT RESTRAIN LAW ENFORCEMENT

The Ninth Circuit, when it rejected Carlos Pena's argument that he was entitled to just compensation, worried that compensating owners like him would discourage law enforcement from taking actions necessary to protect the public. As the court put it, "law enforcement officers (and other government actors) faced with split-second decisions regarding protecting the public and saving lives would need to be constantly attendant to the potential *financial* consequences of their actions."[53]

Former police chief and law enforcement expert Thomas J. Tiderington, however, argues this gets it exactly backward: Officers are trained to do what is necessary without regard to who pays. In amicus briefs urging the Supreme Court to take Carlos Pena's and Amy Hadley's cases, Tiderington, who has more than 35 years of experience training officers at all levels of government, observed that law enforcement trainings and use-of-force policies say nothing about who pays for damage.[54] This is because officers assume it will be the municipality, leaving the planning and execution of an operation up to officers. As he wrote:

> Officers planning a search warrant or a SWAT response weigh officer safety, public safety, suspect apprehension, evidence preservation, and the minimization of collateral harm. They do not pause to ask whose pocket will be lighter when the smoke clears; they just assume it will be added to the municipality's tab. This assumption that the public, rather than individual property owners, will bear those costs is not a deterrent to police work. It is just what makes sense.[55]

Our data support Tiderington's argument. We found hundreds of examples of officers telling property owners how to file claims after damage had occurred, indicating both that these officers understood the costs from their actions could fall on the municipality *and* that they took action anyway. We identified 243 claim records across 73 jurisdictions—nearly 10% of our records—in which claimants reported that police officers referred them to a claim form or, in some cases, even said the government would compensate them.[56] And this is likely an undercount given that claims generally did not include details of conversations with law enforcement.

Police officers in Lincoln, Nebraska, for example, kicked open the wrong door while responding to a domestic disturbance call. According to the property owner's claim, the police department called him to admit the mistake and tell him to "call Risk Management," who then told him to file a claim with the city attorney's office. Another innocent claimant, this one in Winston-Salem, North Carolina, reported that an officer had "advised him to contact Risk Management" after police forced entry into his home in response to a false report of a shooter. The city, the officer told the claimant, "would pay for the damage" to his door. According to the *Statesman,* Eniekenimi Seifere—the Austin, Texas, resident mentioned earlier whose door was mistakenly destroyed—recalled apologetic officers who gave him the contact information for the city's property damage claims administrator.[57]

None of these claimants, however, were compensated. Lincoln denied the broken-door claim, saying there was no "legal basis to establish liability" as "the actions of officers was [sic] appropriate and in accordance with proper emergency procedure." Winston-Salem similarly denied compensation for the broken door there because it "determined that there is no negligence" and police were "preforming [sic] their constitutional duties." And for six years, Austin denied every one of 135 claims it received—despite a requirement that officers explain to property owners how to file a claim.[58] The story was much the same in McKinney, where, as the city's risk manager testified in Vicki Baker's case, police were trained to give her contact information to anyone who might have a property damage claim against the city, even as the city routinely denied claims involving intentional damage.[59]

Officers who tell property owners how to seek compensation for damage may be surprised to learn municipalities often refuse to pay it, as one officer who spoke to the *Statesman* was. He explained that police provide such information "in good faith, thinking that they are helping out the people affected."[60] His experience suggests that the problem may be less that officers are deterred and more that their "good faith" intentions to help property owners often go unfulfilled by municipalities.

Conclusion

When the government takes private property for public use, the Fifth Amendment requires that it pay the owners just compensation. When the government refuses to pay, it imposes burdens on property owners that should rightfully be shared by the public at large. The Fifth Amendment's takings clause was intended to prevent such injustices.

Perhaps the most obvious and best known example is when the government uses eminent domain to take a home to build a road or a school. But the Supreme Court has recognized other government actions as takings that require just compensation, including:

- Flooding acres of privately owned land to improve the navigability of a river.[61]
- Demanding a percentage of farmers' raisin crops to stabilize the market supply.[62]
- Forcing agricultural employers to give labor organizers access to their property.[63]
- Requiring landlords to allow the installation of cable boxes on apartment buildings.[64]

Unfortunately, when it comes to property damage caused by law enforcement acting to protect the public, many governments nationwide deny compensation. As our results show, law enforcement property damage, even when modest, can impose substantial hardship on private individuals. Most victims are innocent bystanders, lacking any relationship to the target of the action that led to their property's damage.

Our results also show that compensating victims of law enforcement property damage is easily accomplished. Paying every claim for this damage, including for damage that might not qualify as a taking, would make virtually no difference to local governments' finances. And in hundreds of cases, officers themselves explained to property owners how to make such claims, indicating they understood that municipalities may have to pay for damage they caused in the course of their duties. As Chief Tiderington put it, officers are "trained to avoid *unnecessary* destruction, but they are not taught to avoid *necessary* destruction to save the government's wallet."[65]

The Supreme Court can, in the two cases IJ has asked it to hear or in future litigation, require governments to pay when actions taken by law enforcement to benefit the public result in private property damage. Our results show that preventing public benefits from becoming private burdens will provide real relief to ordinary people caught in the fray—and do it without straining government budgets.

Appendix A: Detailed Methods

In this appendix, we describe our process for using AI to code the claim records. This includes using human-coded records to train the AI coder and evaluate its performance, as well as generating the final dataset.

TRAINING THE AI CODER

As a first step, we created training and validation datasets with 86 records each. These datasets were manually coded by IJ researchers. Initially, we coded whether the record was relevant to our study. If it was, we coded 18 additional attributes about the record, though as noted in the Methods, we ultimately did not analyze three fields for reasons of reliability.[66]

To ensure the coding was reliable, we employed several quality control procedures. First, we trained the coders on the codebook and coding process. Second, following that training, we required coders to undergo two rounds of practice coding before they began coding in earnest. Third, we gave 20% of records in the training and validation datasets to multiple coders and evaluated their work for consistency. In the event of disagreements, we required coders to meet and come to a consensus about the correct response. In addition, if any field had over 10% initial disagreement on the overlapping portion of the coding, we double-checked the entirety of the coding for that field.

Once the training and validation datasets were human coded and quality checked, we began using the training data to improve the AI coder's performance. As a first step, we established a baseline of performance by using the unedited codebook definition for each question as the model's prompt.[67] We submitted each question for each record to the model as a standalone query.[68] To ensure greater reliability, we ran the model three times for each question for each record and recorded the majority prediction as the final prediction.[69] For example, if the model predicted yes twice and no once, yes was the final prediction. If all three runs produced different predictions, we recorded a null value.

From there, our primary strategy for improvement was automatic prompt optimization, a process in which we iteratively changed the prompts fed to the model and assessed the performance of the changes. We typically started this process by running the original prompt on a small subset of the training data. Then, we used a second AI model ("the reflection model") to analyze the results.[70] The reflection model was fed the PDF record, the correct response from our human coding, the original model's prediction, and sometimes human expert feedback on why errors had occurred. The reflection model would then use this information to generate a new candidate prompt.

If the new candidate prompt performed worse than the original prompt on the small subset of training data, it was discarded. If the new candidate matched or exceeded the prior performance, it moved on to an evaluation with the remainder of the training data. If the candidate was Pareto efficient on the remainder of the training data, it was kept as a "leading candidate."[71] If not, it was discarded.

The above process was repeated for 15 to 30 iterations, with the starting prompt for each iteration randomly selected from the current group of leading candidates.[72] At the end of the process, all leading candidates were evaluated using our validation data. The best performing leading candidate on the validation data was then chosen as the final candidate prompt.[73]

The result of this iterative development process was a set of prompts that substantially elaborated on the initial codebook. For example, the original codebook definition for the determination date question was 171 words. The final prompt was over triple the length at 674 words. (The final prompts we used for every question can be found on IJ's website at https://ij.org/report/public-benefit-private-burden/.)

Typically, the reflection model used several strategies to refine prompts.[74] One of the most common was to clarify failure points. For example, for the determination date, the prediction model would sometimes incorrectly obtain a date from a letter merely acknowledging receipt of a claim instead of the letter providing the government's decision, or determination, on whether to pay the claim. Accordingly, the reflection model proposed prompt revisions that clarified the two types of letters.

Another common strategy the reflection model used was to add examples. For instance, to help differentiate acknowledgment and determination letters, the reflection model added examples of language typically found in each type of letter. For acknowledgement letters, examples included "we have received your claim" and "an investigation will be opened." For determination letters, examples included "we have completed our investigation" and "we are unable to consider your claim."

Finally, the reflection model often added language clarifying key definitions to the beginning of prompts. For example, the reflection model added four core criteria that a record must meet to be relevant to our study. While these four criteria were not explicitly stated in our original codebook, they were an accurate summation of our goals, and their inclusion considerably improved performance.[75]

Once the AI training process was complete, we added a final layer of human review to ensure that all final candidate prompts were still consistent with the original codebook definitions.

As part of this review, we corrected inaccurate language regardless of immediate performance implications. For example, for the claimant type question, the reflection model subtly changed a clause that should have been an "OR" to an "AND." We corrected the clause to "OR" even though the change had not caused any errors in the training or validation data. Fortunately, such issues were rare.

IJ researchers also proposed isolated changes to prompts to fill obvious gaps. For example, they noted that our claim amount prompt did not address whether to count rental car costs following a car accident. We therefore added clarifying language covering such a scenario. We assessed the performance of these suggestions on the validation data and kept the changes so long as they did not negatively affect performance.

For most questions, the foregoing process was sufficient. However, performance was still slightly below our goals for a few questions: a claim's relevance to our study, the claim amount, and the determination date. Accordingly, we coded an additional 150 records for just those three fields using the same human-coding process described above. We used 105 of those records for additional training and the remaining 45 for additional validation. In addition, there were three binary questions (deductible only claims, spike strip damage, and whether law enforcement responded to the wrong address) and several subtypes of relevance for which we had very few examples of the positive class.[76] To supplement our data for focused training, we added 41 human-coded records featuring these types of incidents to just the training data.[77] We then repeated the training process until we were satisfied that no further improvements in performance were likely.

EVALUATING THE AI CODER'S PERFORMANCE

Once the training process was complete, we evaluated the AI coder's performance on records that were completely unseen during the training process. To do this, we had IJ researchers create a gold-standard test dataset of 200 records. This dataset was coded using the same hand-coding and quality control procedures as the training and validation datasets.

Once the test dataset was ready, we ran the same 200 records through the AI coder using the prompts, prediction model, and configurations that we finalized in training.[78] We then compared the AI coder's predictions to the hand coding. For any codes where the initial human coders and the AI coder disagreed, we had a committee of four human experts independently and blindly review the coding decision and decide the final and correct code.[79] In the event the four expert coders did not initially agree, they met and discussed until they reached unanimous agreement.

Ultimately, the AI coder consistently achieved accuracies above 90% and F1 scores above 0.9. For a full range of performance statistics for each field in the study, see Appendix B.

The AI coder performed best on questions that involved retrieving a specific and clear piece of information from somewhere in a record. For example, the incident date was almost always explicitly identified in a claim. As a result, the AI coder's performance was excellent at over 97% accuracy. Similarly, while references to a warrant were not consistently present in the original records, when they were, the AI coder was exceptionally good at finding them. It achieved an F1 score of nearly 0.99 and identified several references to a warrant that the human coders initially missed. Nevertheless, and precisely because references to a warrant were not consistently present in the original records, we opted not to use the field in our final analyses.

Although still solid, the AI coder's performance was not quite as good with subjective interpretations that had no clear right answer—only a best guess based on limited information. For example, on the question about whether the claimant was the reason for the law enforcement action,

there were three options: yes, no, and not enough information to tell. There was not a single instance where the AI coder confused yes and no. Rather, for five records, it made a prediction of yes or no when the human coders determined there was insufficient information, and for another five records, it indicated there was insufficient information when the human coders determined the answer was yes or no.

GENERATING THE FINAL DATASET

Once the evaluation process was complete, we used our final prompts and prediction model to code the remainder of the records. We first ran the model on all the uncoded records to determine whether they were relevant to our study. We then had the model code any relevant records on the remaining variables, plus one additional open-ended question that asked the model for a narrative summary of the record.

We then combined all these predictions with the 490 human-coded records to form a preliminary dataset of roughly 4,500 records, of which roughly 2,700 were relevant.[80]

Before finalizing the dataset, we performed a few quality control adjustments.

First, because each question was a standalone inquiry, there were occasionally inconsistencies between the claim outcome, the paid amount, and the denial reason. For example, for 21 records, the model predicted a claim outcome of either "paid" or "unclear" along with a denial reason. Similarly, for three records, it predicted both an outcome of "denied" and a payment amount. In these 24 instances, we changed the denial reasons or paid amounts to null values to ensure consistency with the claim outcome.

Second, 28 jurisdictions that did not provide determination documents with their records sent separate spreadsheets with determination information for each decided claim. For 474 relevant claims, or 17% of our final dataset, we replaced the model's predictions with the claim outcomes and payment amounts from those spreadsheets. As a result, the claim outcome and payment amount fields are likely more accurate than reported in Appendix B.

Third, we flagged and removed 12 duplicate claims.[81]

Fourth, to guard against large errors influencing the claim amount and paid amount fields, we flagged outliers for further manual review and correction.[82] For the claim amount, we flagged 57 records as outliers and ultimately corrected 24 of them. For the paid amount, we flagged 24 records and corrected two of them.

Finally, we attempted to reconcile occasional inconsistencies between the claim amount and paid amount fields. Most commonly, either the paid amount exceeded the claim amount, or there was a paid amount but no claim amount. After some exploratory review, we found that there were generally good reasons for these inconsistencies. For example, the claim amount was often an estimate or unknown when the claim was initially filed. However, we noticed a prevalence of errors in higher-dollar claims, as

these claims often included bodily injuries in addition to property damage. Accordingly, we opted to manually review every inconsistency that involved either a claim amount or a paid amount over $10,000. This review resulted in 16 corrections to the claim amount, 20 corrections to the paid amount, and 21 records becoming irrelevant as they did not include property damage. Our final dataset contains 4,466 records of which 2,752 are relevant property damage claims.

Appendix B: Performance Statistics

This appendix provides a comprehensive overview of our AI coder's performance on every field in our study. For the relevance field, we calculated performance statistics based on the entire 200-claim test dataset. For all other fields, we calculated performance statistics based on the 114 relevant records in the test dataset. We did not use the test data at any point during the process of developing the AI coder.

To measure performance, we used the following statistics:

- **Accuracy**: Accuracy measures the number of correct predictions out of the total number of predictions for a field. For example, if the AI coder predicts a field 100 times and 93 predictions are correct, then the accuracy is 93%. Accuracy is our primary metric for evaluating date, numeric, and categorical fields. Although we report accuracy for all predicted fields, other metrics—precision, recall, and F1 score—are better indicators of performance for binary (yes/no) fields.
- **Precision**: In a nutshell, precision measures how good the AI coder is at avoiding false positives. For example, if the AI coder predicts that 20 claims involve a pursuit, but only 18 actually do (meaning two were false positives), then the precision is 0.9 (18/20). The maximum value is 1. We report precision only for binary fields and for the individual response options for questions where multiple categorical responses are possible (i.e., "select all that apply" questions).
- **Recall**: Put simply, recall measures how good the AI coder is at picking out the field in question. For example, if there are 25 true tactical raids and the AI coder correctly identifies 23 of them, then recall is 0.92 (23/25). The maximum value is 1. We report recall only for binary fields and for the individual response options for questions where multiple categorical responses are possible.
- **F1 Score**: The F1 score is a widely used performance metric that combines precision and recall into a single statistic by taking their harmonic mean. A harmonic mean is a mean that penalizes divergence between the values being averaged. For example, the harmonic mean of 0.8 and 0.8 is 0.8, but the harmonic mean of 0.6 and 1 is 0.75, because 0.6 and 1 have greater divergence than 0.8 and 0.8. A high F1 score indicates that both precision and recall performed well. The maximum value is 1.
- **Confusion Matrix**: We also provide confusion matrices, which offer additional context for the AI coder's performance. Although not technically a statistic, a confusion matrix compares predictions to the true values by putting them into a table. In this table, the rows contain the ground truth values, while the columns contain the predictions. For example, if the ground truth for a claim was "B" but the AI coder predicted "C," the claim would be

tabulated in the row for “B” and in the column for “C.” In our matrices, values shaded in green are correct predictions—that is, predictions that matched the true value—while values shaded orange are incorrect predictions. Gray is used for cells with no true or predicted values. Date and amount fields do not have confusion matrices, while “select all that apply” fields have precision, recall, and F1 scores calculated for the individual response options.

<table>
<tr><th></th><th>Field</th><th>Type</th><th>Accuracy</th><th>Precision</th><th>Recall</th><th>F1 Score</th><th>Confusion Matrix
Ground truth: rows
Prediction: columns</th></tr>
<tr><td>Relevance</td><td>Relevance</td><td>Binary</td><td>92.5%</td><td>0.884</td><td>1.000</td><td>0.938</td><td>
<table>
<tr><th></th><th>0</th><th>1</th></tr>
<tr><td>0</td><td>71</td><td>15</td></tr>
<tr><td>1</td><td>0</td><td>114</td></tr>
</table>
</td></tr>
<tr><td rowspan="5">Claim Basics</td><td>Claim amount[a]</td><td>Numeric</td><td>91.2%</td><td>--</td><td>--</td><td>--</td><td>--</td></tr>
<tr><td>Deductible only</td><td>Binary</td><td>100%</td><td>N/A[b]</td><td>N/A[b]</td><td>N/A[b]</td><td>
<table>
<tr><th></th><th>0</th><th>1</th></tr>
<tr><td>0</td><td>114</td><td>0</td></tr>
<tr><td>1</td><td>0</td><td>0</td></tr>
</table>
</td></tr>
<tr><td>Incident date</td><td>Date</td><td>97.4%</td><td>--</td><td>--</td><td>--</td><td>--</td></tr>
<tr><td>Claim date</td><td>Date</td><td>93.9%</td><td>--</td><td>--</td><td>--</td><td>--</td></tr>
<tr><td>Type of damage</td><td>Select all that apply</td><td>94.7%</td><td colspan="3">
<table>
<tr><th></th><th>N</th><th>Precision</th><th>Recall</th><th>F1</th></tr>
<tr><td>A</td><td>73</td><td>0.99</td><td>0.99</td><td>0.99</td></tr>
<tr><td>B</td><td>7</td><td>1.00</td><td>1.00</td><td>1.00</td></tr>
<tr><td>C</td><td>1</td><td>1.00</td><td>1.00</td><td>1.00</td></tr>
<tr><td>D</td><td>33</td><td>1.00</td><td>1.00</td><td>1.00</td></tr>
<tr><td>E</td><td>10</td><td>0.63</td><td>1.00</td><td>0.77</td></tr>
</table>
As individual records can have multiple types of damage, we do not include a confusion matrix and instead present the performance statistics for each response category.
</td><td>--</td></tr>
<tr><td rowspan="2">Claimant Information</td><td>Claimant type</td><td>Categorical</td><td>89.5%</td><td>--</td><td>--</td><td>--</td><td>
<table>
<tr><th></th><th>A</th><th>B</th><th>C</th><th>D/E</th><th>F</th></tr>
<tr><td>A</td><td>68</td><td>0</td><td>3</td><td>0</td><td>6</td></tr>
<tr><td>B</td><td>0</td><td>2</td><td>0</td><td>1</td><td>0</td></tr>
<tr><td>C</td><td>0</td><td>0</td><td>2</td><td>1</td><td>1</td></tr>
<tr><td>D/E</td><td>0</td><td>0</td><td>0</td><td>15</td><td>0</td></tr>
<tr><td>F</td><td>0</td><td>0</td><td>0</td><td>0</td><td>15</td></tr>
</table>
</td></tr>
<tr><td>Claimant was target</td><td>Categorical</td><td>91.2%</td><td>--</td><td>--</td><td>--</td><td>
<table>
<tr><th></th><th>A</th><th>B</th><th>C</th></tr>
<tr><td>A</td><td>8</td><td>0</td><td>0</td></tr>
<tr><td>B</td><td>0</td><td>82</td><td>5</td></tr>
<tr><td>C</td><td>2</td><td>3</td><td>14</td></tr>
</table>
</td></tr>
</table>

<table>
<tr><th></th><th>Field</th><th>Type</th><th>Accuracy</th><th>Precision</th><th>Recall</th><th>F1 Score</th><th>Confusion Matrix
Ground truth: rows
Prediction: columns</th></tr>
<tr><td>Claimant Information</td><td>Relationship to target</td><td>Categorical</td><td>89.5%</td><td>--</td><td>--</td><td>--</td><td>
<table>
<tr><th></th><th>A</th><th>B</th><th>C/D</th><th>E/F</th><th>G</th><th>N/A</th></tr>
<tr><td>A</td><td>49</td><td>2</td><td>0</td><td>0</td><td>3</td><td>0</td></tr>
<tr><td>B</td><td>0</td><td>17</td><td>0</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td>C/D</td><td>0</td><td>0</td><td>6</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td>E/F</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td>G</td><td>1</td><td>0</td><td>3</td><td>1</td><td>22</td><td>2</td></tr>
<tr><td>N/A</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>8</td></tr>
</table>
</td></tr>
<tr><td rowspan="4">Outcome Information</td><td>Claim outcome[a]</td><td>Categorical</td><td>95.6%</td><td>--</td><td>--</td><td>--</td><td>
<table>
<tr><th></th><th>A</th><th>B</th><th>C</th></tr>
<tr><td>A</td><td>38</td><td>0</td><td>0</td></tr>
<tr><td>B</td><td>1</td><td>29</td><td>0</td></tr>
<tr><td>C</td><td>1</td><td>3</td><td>42</td></tr>
</table>
</td></tr>
<tr><td>Paid amount[a]</td><td>Numeric</td><td>93.0%</td><td>--</td><td>--</td><td>--</td><td>--</td></tr>
<tr><td>Reason for denial</td><td>Select all that apply</td><td>90.4%</td><td colspan="3">
<table>
<tr><th></th><th>N</th><th>Precision</th><th>Recall</th><th>F1</th></tr>
<tr><td>A</td><td>0</td><td>0.00</td><td>N/A[b]</td><td>N/A[b]</td></tr>
<tr><td>B</td><td>1</td><td>0.50</td><td>1.00</td><td>0.67</td></tr>
<tr><td>C</td><td>0</td><td>N/A[b]</td><td>N/A[b]</td><td>N/A[b]</td></tr>
<tr><td>D</td><td>0</td><td>N/A[b]</td><td>N/A[b]</td><td>N/A[b]</td></tr>
<tr><td>E</td><td>26</td><td>0.90</td><td>1.00</td><td>0.95</td></tr>
<tr><td>F</td><td>1</td><td>0.33</td><td>1.00</td><td>0.50</td></tr>
<tr><td>G</td><td>5</td><td>0.50</td><td>0.80</td><td>0.62</td></tr>
<tr><td>H</td><td>1</td><td>1.00</td><td>1.00</td><td>1.00</td></tr>
<tr><td>I</td><td>12</td><td>0.92</td><td>1</td><td>0.96</td></tr>
<tr><td>J</td><td>11</td><td>1.00</td><td>0.82</td><td>0.90</td></tr>
</table>
As individual records can have multiple reasons for denial, we do not include a confusion matrix and instead present the performance statistics for each response category.
</td><td>--</td></tr>
<tr><td>Determination date</td><td>Date</td><td>92.1%</td><td>--</td><td>--</td><td>--</td><td>--</td></tr>
<tr><td rowspan="2">Incident Circumstances</td><td>Pursuit</td><td>Binary</td><td>97.4%</td><td>0.864</td><td>1.000</td><td>0.927</td><td>
<table>
<tr><th></th><th>0</th><th>1</th></tr>
<tr><td>0</td><td>92</td><td>3</td></tr>
<tr><td>1</td><td>0</td><td>19</td></tr>
</table>
</td></tr>
<tr><td>Tactical raid</td><td>Binary</td><td>99.1%</td><td>0.944</td><td>1.000</td><td>0.971</td><td>
<table>
<tr><th></th><th>0</th><th>1</th></tr>
<tr><td>0</td><td>96</td><td>1</td></tr>
<tr><td>1</td><td>0</td><td>17</td></tr>
</table>
</td></tr>
</table>

	Field	Type	Accuracy	Precision	Recall	F1 Score	Confusion Matrix *Ground truth: rows* *Prediction: columns*
Incident Circumstances	Warrant[c]	Binary	99.1%	0.972	1.000	0.986	0: 78, 1 / 1: 0, 35
	Wrong address[d]	Binary	97.4%	0.500	1.000	0.667	0: 108, 3 / 1: 0, 3
	Spike strips	Binary	100%	1.000	1.000	1.000	0: 112, 0 / 1: 0, 2
	Death[d]	Binary	97.4%	0.571	1.000	0.727	0: 107, 3 / 1: 0, 4

Note: See codebook in Appendix C for the definitions of each lettered response category.

[a] We could not calculate this metric due to insufficient data.

[b] We made additional manual corrections for these fields. See Appendix A for details.

[c] Though the AI coder achieved high accuracy as compared to human coders on the warrant field, our researchers' review made clear that the underlying records are not fully reliable, as claimants may not know or indicate on a claim form whether a warrant was involved and associated records may also not indicate the presence of a warrant. Therefore, we elected not to analyze this field.

[d] Due to both their rarity and the poor precision of the AI coder, we opted not to analyze these fields.

Appendix C: Codebook

This appendix contains definitions for the fields we manually coded and subsequently predicted with our AI coder. We provided it to all human coders to use as a guide during the coding process. We have made minor edits for clarity

Global note for coders: When a question permits more than one response in a multiple-choice question, place a comma (and no spaces after it) in between your responses if you choose more than one response.

1. **Relevance: Is the record relevant?**

   This question will likely take the longest time to get up to speed on, though after a while, it should be one of the easier questions to answer. You only need to code a single letter for this (L if it is relevant, a different letter if it is irrelevant). If you believe the record is irrelevant for a reason other than any listed below, choose R and include a note about your reasoning in the notes column.

   A. The record does not include any evidence of a claim being filed. For example, there is no claim form, there are no other documents submitted by a claimant indicating a claim was made, or no government document of any kind indicating a claim, claim amount, or information about the claim. NOTE: Even if there is a determination, the record is irrelevant if there is no record of a claim.
   B. The damage was not caused by law enforcement.
   C. No property damage is claimed. For instance, the claimant only alleged bodily injury.
   D. The claimant is seeking compensation that was already paid to a different party. Duplicate claims would fall under this category.
   E. The claimant does not own the property that was damaged. NOTE: Do not select this if a claim is filed on behalf of the person who owns the damaged property.
   F. The claim involved property that was lost or forgotten during a roadside stop, while the claimant was being booked into jail, or while the claimant was being transported to a police station or other law enforcement facility. Records in which property stored as evidence was lost, destroyed, or damaged are also irrelevant.
   G. The damage was caused by an officer who was in traffic like any other motorist. In other words, this was a fender-bender type incident *not* caused while the officer was pursuing a suspect or driving to a call.
   H. The claimant's property was damaged while the claimant was attacking an officer.

I. The damage was caused during a wellness check. However, there are a few types of wellness checks that would still be considered relevant: (1) If the wellness check was performed on the wrong address, do not select this option; (2) if the wellness check was performed in response to a domestic violence call, do not select this option; (3) if police were checking on residents who may have been alarmed or harmed by a nearby action such as a shootout, do not select this option; (4) if police were performing a wellness check and had reason to believe violence to self or others was imminent, do not select this option.

J. A *different* law enforcement agency was responsible for the damage.

K. None of the above or below apply—the record is relevant.

L. The damage was related to a vehicle being towed or was related to the cost of the tow.

M. The claimant reported their property was stolen (e.g., auto theft) but then claimed law enforcement was responsible for the theft or any damages caused by the thieves.

N. There is not enough information to know what happened and/or who caused the damage.

O. The incident involved a seizure of property (for evidence or forfeiture) as opposed to damaged property.

P. The damage was caused by gates, barriers, or doors at a police facility.

Q. Other: Please record the reason in the notes column.

Expected output: The letter K if the record is relevant; otherwise, the letter of the exclusion criterion that applies to the record.

2. **Claim amount: What is the amount of property damage claimed?** Please record the exact amount—do not round.

For this, code the amount claimed based on the following priority ranking: (1) the amount of property damages claimed that is written or typed on the claim form itself; (2) an amount indicated on the government's summary or abstract of the claim record—this may sometimes appear as "amount demanded" or "initial amount demanded"; (3) an invoice for cost of repairs/replacement; (4) a document containing an estimate of how much the repair/replacement will cost. If you do not see any of the aforementioned four items, leave the amount blank. If the amount is presented as a range (e.g., a high and a low estimate), record the lower end of the range. If multiple amounts are provided (e.g., for multiple damaged items), record each value separated by a + sign (no spaces). When recording amounts for which the claimant did not include cents (e.g., $100 or $5 or $1,000 as opposed to $100.59 or $25.99, etc.) just record the whole number (e.g., 100 or 5 or 1000). But if

cents are included in the amount the claimant wrote, use a decimal point to report the exact amount they used.

We recommend double-checking the amount you wrote to make sure it matches what appears in the record.

Expected output: A numeric dollar amount.

3. **Deductible only: Does the record indicate that the claimant is requesting reimbursement only for an insurance deductible?**
   This appears to be rare, but it does happen occasionally. Typically, a deductible amount will be a round number such as $1,000. Further, if an insurance company is filing a subrogation claim that includes both the damages it already paid to its insured customer and the deductible, do not answer yes to this question.

   Expected output: Yes or No.

4. **Incident date: What was the date of the incident as it appears in the record?**
   Use the date recorded by the claimant on the claim form. If there is no "date of loss" or incident date on the claim form, use the earliest date appearing in any government documents that include the date the incident occurred (e.g., a police report or a determination letter or email correspondence that includes the date of the incident). The terminology typically used for the incident date includes "date of loss," "date of incident," "date of occurrence," "event date," etc.

   Expected output: A date, written as MM/DD/YYYY.

5. **Claim date: When was the claim submitted?**
   For this, prioritize the date recorded by the claimant on the claim form. This is typically located next to the claimant's signature toward the end of the form, but not always. If there is no date on the claim form, use the earliest date appearing in any government documents that include the date the government received the submission.

   Prioritize the date recorded by the claimant on the claim form here. Why? Because we cannot always count on the records containing a stamp or other feature that includes the submission date (or the date it was received by the government). If there is no date recorded by the claimant, then look for a

"claim received on" or "claim submitted on" or "report date" elsewhere in the record. Sometimes, a government stamp will be visible on the claim form that includes the day the claim was received. But remember, if the claimant included a date next to their signature, use that date.

Expected output: A date, written as MM/DD/YYYY.

6. **Claimant type: What best describes the person submitting the claim?**
   These will typically be people filing on behalf of themselves, but we also see records submitted by lawyers on behalf of a client (B) or insurance companies on behalf of an insured customer (D) or an insurance company filing a subrogation claim in which it is attempting to be reimbursed by the government for damages for which it already compensated an insured customer (E).

   A. An individual filing on behalf of themselves.
   B. A lawyer filing on behalf of a client.
   C. A non-lawyer filing on behalf of someone else.
   D. An insurance company filing on behalf of an insured customer.
   E. An insurance company filing on behalf of itself. This may be called a subrogation claim, in which it is seeking compensation from an at-fault party.
   F. A company owner or a company name (like a doing-business-as name) or a company representative (like a CEO or property manager) filing on behalf of the company.

   Expected output: The letter representing the most appropriate single response from the above list of options A through F.

7. **Claim outcome: How did the government respond to the claim?**
   Only code this (and all other fields) based on what you see in the PDF record. Do not code information from any separate determination records associated with the government responsible for the record in question (that is, documents not included in the PDF). Why? Because the AI tool will not see those separate determination records. If you code a field based on a separate determination document, the field will be marked as incorrect because the AI tool will respond with C here.

   A. The government denied the claim.
   B. The government paid/settled the claim. NOTE: The terms "paid" and "settled" are used interchangeably.

C. The record is silent as to the government's determination on the claim.

Expected output: The letter representing the most appropriate single response from the above list of options A through C.

8. **Paid amount: If you answered B to Question 7 (i.e., the government paid/settled the claim), exactly how much did the government pay?** Please record the exact amount—do not round. Two key notes: First, if the claim was denied (that is, you answered A for the previous question), do not enter any value here, even zero. Second, it is not usual for the government to make multiple payments, but it happens occasionally. If this is the case, enter the amounts separated by + signs.

   Expected output: A numeric dollar amount.

9. **Reason for denial: If you answered A to Question 7 (i.e., the government denied the claim), why was the claim denied?**
   The rationale, if provided, will typically be found in a determination letter. But it can also be found elsewhere in the record, such as in correspondence among government staff when discussing the decision on the claim. The most common denial codes are likely to be E, I, and J. If I was selected below, E should be automatically selected as well. Please code any denial reason present in the record, with each reason separated by a comma (e.g., E,I).

   A. Incorrect submission format or incorrect filing process. This includes things like an absence of required receipts, estimates, etc.
   B. Claim was filed after the deadline for submissions.
   C. State forbids subrogation claims.
   D. State forbids indemnification claims.
   E. Government reported some version of "the officer was performing a lawful action within the course and scope of their employment" or "the officer was acting under color of law" or "the officer was operating under statutory authority" or "the officer was following all policies and procedures" or some other legalistic phrase synonymous with these phrases.
   F. The government noted that the claimant was the target of the action.
   G. The government noted that the suspect/target of the action was responsible for the damage and/or the claimant should seek reimbursement from the suspect/target of the action.

H. The government noted that the claimant should instead seek reimbursement from a victim's compensation fund.
I. The government noted that law enforcement had a warrant to conduct the action.
J. No rationale provided.

Expected output: A list of all denial reasons appearing in the record in alphabetical order, with each reason separated by a comma.

10. Determination date: When was the determination made?

You will sometimes see a number of different dates that could correspond to the determination date. For this, code the date based on the following priority ranking: (1) a letter that specifically states the day that the determination was made (e.g., "the City Council rejected your claim on 7/7/25" or language about how the claim was denied/rejected/approved by operation of law on a specific date); (2) the date at the top of the determination letter indicating when it was written and/or sent to the claimant, if the letter does not contain specific information about when the claim was rejected or paid/settled; (3) a date found in a claim abstract or summary that mentions the day that the case was decided/resolved/closed/etc.; (4) the date on a settlement release form (which might contain language such as "release of all claims"); (5) the date on the check or check stub sent to the claimant and included in the record.

Expected output: A date, written as MM/DD/YYYY.

11. Type of damage: What type(s) of property damage was claimed?

Three key notes: First, most of these will be damage to residential property or vehicles, but there are sometimes other types of property damaged. Please note that occasionally claimants will claim damage to their yard, grass, or other outdoor area. Unless it is explicitly clear that such land was commercial/industrial or agricultural, please respond with A. Second, claimants occasionally do claim *personal* property damage, which includes non-real estate/non-vehicle items like TVs, books, beds, etc. So keep a watchful eye for such claims. Personal property here is property that is *not* permanently affixed to the structure, so a picture that was simply hung on a nail is personal property, but a mirror that is glued or bolted to a wall is real property and would simply be coded as residential property damage (see https://www.rocketmortgage.com/learn/fixture-real-estate for more on how to decide whether something is personal or real property). Third, trailer homes should be treated as personal property unless there is clear evidence that the property owner owns the land the trailer is on and the trailer is permanently affixed to the land. Fourth, claimants sometimes

claim damages to multiple types of property (e.g., a home and a car or a home and personal property like a TV within the home). Please code any type of property they claimed damages for, with each type separated by a comma (e.g., A,E).

A. Damage to residential property. This includes apartments and home-based Airbnbs but not hotels or vehicles.
B. Damage to commercial or industrial property.
C. Damage to agricultural property.
D. Damage to a vehicle.
E. Damage to personal property, such as a television.

Expected output: A list of all types of property damage claimed in the record in alphabetical order, with each type separated by a comma.

12. Pursuit: Was the damage due to a pursuit by law enforcement?

Here, you should choose yes if a record includes any active pursuit of someone who is fleeing from law enforcement. Examples of pursuits include (1) an officer pursuing a target in a car; (2) an officer running through a yard trying to catch a target; (3) incidents in which spike strips were deployed to stop a target's car from moving; (4) K-9s chasing after a target; (5) an active tactical raid of a building *if* law enforcement had to chase someone to that building; (6) an officer breaking into a building *after* chasing someone there. Examples of incidents that should not be coded as pursuits: An officer driving with sirens and lights to get to a house for a wellness check; an officer breaking into a building to serve a search or arrest warrant (unless a suspect flees after entry and the police chase them); a SWAT or other tactical raid that never required law enforcement to chase the target of the raid.

The key here is that for it to be considered a pursuit, the target of the law enforcement action must, at some point, have been actively fleeing from law enforcement, who were (again, at some point) chasing them. The use of spike strips counts because it is an attempt to stop a fleeing target. Finally, a patrol car with lights and sirens on is not necessarily associated with a pursuit. You must have evidence in the record that the officer was chasing a fleeing suspect at some point during the event.

Expected output: Yes or No.

13. **Tactical raid: Was the damage caused during a SWAT/tactical raid?**

    Answer yes here if it is certain or very likely that the incident involved a tactical raid of any type. Keep in mind that SWAT teams are also known as Community Emergency Response Team (CERT), Emergency Response Team (ERT), Special Operations Team (SOT), Rapid Response Team (RRT), etc. Any mention of a tactical team (including bomb squads) or a raid should lead you to answer yes to this question.

    The use of specialized tools like chemical munitions (e.g., tear gas), Bearcats (a tactical vehicle), or flash-bangs is usually a very good sign that a tactical raid occurred, particularly if the event included a barricaded individual. The use of a battering ram, however, should not lead you to code this as a SWAT/tactical raid unless other elements of a raid are evident in the record.

    Expected output: Yes or No.

14. **Warrant: Was any type of warrant served during or otherwise associated with the incident?**

    (Note: We did not use this field in our final analysis.)

    For warrant service, only respond yes if a warrant is mentioned somewhere in the record by the claimant or the government. Do not assume a warrant was associated with the incident unless the word "warrant" explicitly appears somewhere in the record. For example, if the record notes that the target of the action was wanted on a warrant, you should answer yes.

    Expected output: Yes or No.

15. **Wrong address: Was the damage caused during an action performed at the wrong address?**

    (Note: We did not use this field in our final analysis.)

    To select yes, you must have evidence in the record that the address was indeed incorrect. Here, you can select yes in the following situations: (1) the government says the police went to the wrong place; (2) the claimant says the police went to the wrong place and the government ultimately pays or settles the property damage claim; or (3) there is enough context in the record to be confident that police indeed went to the wrong address. If there is no explicit or implicit evidence anywhere in the record that a wrong address was involved, select no.

    Expected output: Yes or No.

16. **Spike strips: Did the damage involve spike strips set out for a different driver?**
    These are also known as stop strips, stop spikes, etc. Here, we will assume the claimant was not the target of the spike strip action unless the government says they were. This assumption is based on observing many spike strip claims in our records. Governments vary widely in terms of paying spike strip claims—some pay all of them, while others pay none or some—so you cannot use a denied claim to assume that the spike strip was intended for the claimant.

    Expected output: Yes or No.

17. **Death: Did *any* person or animal die in the incident or as a result of the incident?** *(Note: We did not use this field in our final analysis.)*
    This will be quite rare.

    Expected output: Yes or No.

18. **Claimant was target: Given the available information and context of the record, was the claimant the reason for the law enforcement action?**
    Sometimes the record will not have enough information to answer one way or another on this question. In those situations, respond with C.

    A. Yes, the claimant was definitely or likely the reason for the action.
    B. No, the claimant was definitely or likely not the reason for the action.
    C. There is not enough clear information in the record to know whether the claimant likely was or likely was not the reason for the action.

    Expected output: The letter representing the most appropriate single response from the above list of options A through C.

19. **Relationship to target: If you answered B or C to Question 18, given the information and context of the record, did the claimant have any sort of relationship with the person who *was* the reason for the action?**
    This will commonly be A when vehicle damage is involved (e.g., unintended spike strip victims or someone hit by a police car that was chasing after someone else). If the damage was to a vehicle during some sort of vehicle-caused accident, you can safely choose A unless there is something in

the record that clearly says otherwise. The option B is common when law enforcement damages an apartment.

A. The claimant has no relationship with the person who was the reason for the action.
B. The claimant is the landlord of the person who was the reason for the action or otherwise contracts their property out to that person. This includes situations where the claimant is the owner or representative of a hotel in which targets of the police action were staying.
C. The claimant is a parent or guardian of the person who was the reason for the action.
D. The claimant has some other familial relationship with the person who was the reason for the action (e.g., spouse, cousin, aunt, uncle, etc.).
E. The claimant has a friendly or romantic relationship with the person who was the reason for the action (e.g., friends or significant others).
F. The claimant has some other type of relationship with the person who was the reason for the action (e.g., coworkers or roommates).
G. The record is silent about whether the claimant has any type of relationship with the person who was the reason for the action.

Expected output: The letter representing the most appropriate single response from the above list of options A through G. If you did not answer B or C to Question 18, simply reply with n/a.

Endnotes

1 Complaint, *Pena* v. *City of Los Angeles*, No. 2:23-cv-05821 (C.D. Cal. July 19, 2023), https://ij.org/wp-content/uploads/2023/07/LA-SWAT-Complaint-FILE-STAMPED-07.19.2023.pdf [hereinafter *Pena* Complaint].

2 Complaint, *Baker* v. *City of McKinney*, No. 4:21-cv-00176 (E.D. Tex. Mar. 3, 2021), https://ij.org/wp-content/uploads/2021/03/ECF-1-Complaint-for-Compensatory-Damages-FILE-STAMPED-03.03.21.pdf [hereinafter *Baker* Complaint].

3 *Pena* Complaint, *supra* note 1, at ¶ 36; *Baker* Complaint, *supra* note 2, at ¶ 27; 5 New Appleman on Insurance Law Library Edition § 43.02[8] (Jeffrey E. Thomas & Susan Lyons eds., 2026 update).

4 Grumet, B. (2026a, May 7). 'I was shaking': Why Austin residents pay when APD kicks in the wrong door. *Austin American-Statesman*. https://www.statesman.com/opinion/damaged-for-good/article/austin-police-wrong-door-damages-22160487.php. The *Statesman*'s full "Damaged for Good" series can be found here: https://www.statesman.com/opinion/damaged-for-good/

5 In addition to the larger scope, our study differs from the *Statesman*'s analysis in other important ways. For instance, we looked at claims for damage to all types of physical property and excluded claims filed by anyone other than property owners or their agents. The *Statesman*, in contrast, limited its analysis to claims for damage to buildings and included claims filed by renters.

6 Bennett, K. (2026, February 4). *Bankrate's 2026 annual emergency savings report*. Bankrate. https://www.bankrate.com/banking/savings/emergency-savings-report/

7 Petition for Writ of Certiorari, *Baker* v. *City of McKinney*, 145 S. Ct. 11 (2024) (No. 23-1363), https://www.supremecourt.gov/DocketPDF/23/23-1363/315950/20240628093743508_1%20-%20Cert%20Petition_FINAL.pdf

8 *Baker* v. *City of McKinney*, 145 S. Ct. 11, 12 (2024) (Sotomayor, J., statement respecting the denial of certiorari).

9 *Id.* (quoting *Armstrong* v. *United States*, 364 U.S. 40, 49 (1960)).

10 *Id.* at 13. The Supreme Court's decision not to hear Vicki's case put an end to her claim under the U.S. Constitution, but Vicki still had a state constitutional claim. In June 2025, a federal judge ruled that Vicki is entitled to damages, plus interest, under the Texas Constitution. *Baker* v. *City of McKinney*, No. 4:21-cv-00176, 2025 WL 1492949 (E.D. Tex. June 5, 2025). And in May 2026, the Fifth Circuit upheld that ruling. *Baker* v. *City of McKinney*, No. 25-40396, 2026 WL 1453172 (5th Cir. May 22, 2026).

11 Petition for Writ of Certiorari at 2, *Pena* v. *City of Los Angeles*, No. 25-1163 (U.S. Apr. 6, 2026), https://www.supremecourt.gov/DocketPDF/25/25-1163/403706/20260406114941744_Pena%20v.%20City%20of%20Los%20Angeles%20-%20Petition%20for%20Writ%20of%20Certiorari.pdf [hereinafter *Pena* Petition]; Petition for Writ of Certiorari, *Hadley* v. *City of South Bend*, No. 25-1158 (U.S. Apr. 6, 2026), https://www.supremecourt.gov/DocketPDF/25/25-1158/403699/20260406112159213_Hadley%20v.%20City%20of%20South%20Bend%20-%20Petition%20for%20Writ%20of%20Certiorari.pdf [hereinafter *Hadley* Petition].

12 The largest 20% of agencies with SWAT teams had a minimum of 90 sworn officers. Our source for whether an agency used a SWAT team was the Bureau of Justice Statistics' Census of State and Local Law Enforcement Agencies. See https://bjs.ojp.gov/data-collection/census-state-and-local-law-enforcement-agencies-csllea

13 We selected 467 jurisdictions because a 60% response rate would have yielded 280 jurisdictions and a margin of error of +/- 5% at the 95% confidence interval.

14 Often, the spreadsheets contained only the claimant's name, the incident date, and a one- or two-sentence description. This was generally not enough information to code most fields in our study.

15 The 222 jurisdictions come from 44 states and the District of Columbia. The only states not represented are Alabama, Maine, Nevada, South Dakota, Vermont, and Wyoming. Our random sample of 467 agencies included agencies from Alabama, Maine, and Wyoming, but none of their parent jurisdictions responded to our requests. Nevada and South Dakota had agencies that were eligible for inclusion, but none of them were sampled due to random chance. Vermont had no eligible agencies.

16 In addition to the variables shown in Table 1, we coded claims on whether (1) the damage was caused by law enforcement responding to the wrong address, (2) a person or animal died during the incident, and (3) there was a warrant associated with the incident. Both wrong addresses and deaths were rare. Moreover, the AI coder tended to say they were present when they were not: The F1 scores for wrong address and death were 0.667 and 0.727, respectively. As for warrants, although the AI coder achieved an excellent F1 score of 0.986, our researchers' review made clear that the underlying records are not fully reliable. This is because claimants may not know or indicate on a claim form whether a warrant was involved and associated records may not indicate the presence of a warrant either. Therefore, we did not use the three fields in our final analysis. We also omitted them from the final dataset.

17 See, e.g., Sargeant, H., Izzidien, A., & Steffek, F. (2026). Topic classification of case law using a large language model and a new taxonomy for UK law: AI insights into summary judgment. *Artificial Intelligence and Law*, *34*(2), 443–491. https://doi.org/10.1007/s10506-025-09434-0 (88.2% accuracy on multiclass topic sorting); Coan, T. G., Malla, R., Nanko, M. O., Kattrup, W., Roberts, J. T., Cook, J., & Boussalis, C. (2026). Large language model reveals an increase in climate contrarian speech in the United States Congress. *Communications Sustainability*, *1*. https://doi.org/10.1038/s44458-025-00029-z (best model had an F1 score of 0.852 for topic classification); Izzidien, A., Sargeant, H., & Steffek, F. (2024). LLM vs. lawyers: Identifying a subset of summary judgments in a large UK case law dataset. *arXiv*. https://doi.org/10.48550/arXiv.2403.04791 (F1 score was 0.94 for classification of whether a case involved summary judgment); Sapiro-Gheiler, E., & Kastellec, J. P. (2026). *Appealing to large-language-model-as-judge: A comprehensive machine-coded database for the U.S. Courts of Appeals* (Working Paper). https://jkastellec.scholar.princeton.edu/sites/g/files/toruqf3871/files/documents/sapiro_gheiler_kastellec_Large-Language-Model-as-Judge_june_2026.pdf (accuracies ranging from 54% to 93% for classification of various attributes of appellate decisions).

18 The F1 score measures how good the model is at avoiding false positives while also picking out true positives. More technically, it is the harmonic mean of precision and recall. See Appendix B for further details.

19 To infer from our sample to the target population, we used a cell-based post-stratification weighting procedure. This involved breaking our sample into groups, or "cells," where each cell contained similar jurisdictions as defined by one or more relevant variables. We then weighted each cell based on its proportion in our sample and its proportion in the country. If a cell was underrepresented in our sample relative to the country, it received a higher weight. If it was overrepresented, it received a lower weight. To determine the characteristics that defined our cells, we ran several logistic regressions and bivariate tests that explored whether certain variables were associated with an increased or decreased likelihood of receiving a complete public records request response from a jurisdiction. Based on these analyses, the jurisdiction's population, the number of sworn officers, and the agency type (police departments versus sheriff's offices) were not significantly related to responsiveness. However, the agency's location in the country (as defined by its census division) was strongly related to responsiveness. (Only 18% of agencies in the East South Central region responded, compared to roughly 70% in the Mountain and Pacific regions.) We therefore used the census division as the defining characteristic of our cells, with agencies from the lower-responding regions receiving more weight to offset their regions' lower response rates.

20 The technical criteria for each broad claim type are below. The codebook in Appendix C provides definitions of property types, tactical raids, pursuits, spike strip incidents, and claimant types.
Residential damage from tactical raid: (1) claim involved residential property damage; AND (2) claim involved a tactical raid. This category can include damage from pursuits that developed into tactical raids, as well as damage to other types of property if residential damage was also present.
Residential damage from pursuit: (1) claim involved only residential property damage OR only residential *and* personal property damage; AND (2) claim involved a pursuit; AND (3) claim did not involve a tactical raid; AND (4) claim did not involve spike strips.
General residential damage (no pursuit or tactical raid): (1) claim involved only residential property damage OR only residential *and* personal property damage; AND (2) claim did not involve a tactical raid; AND (3) claim did not involve a pursuit.
Vehicle damage, no spike strips: (1) claim involved only vehicle property damage AND did not involve spike strips or a tactical raid; OR (2) claim involved a pursuit with vehicle property damage (and possibly damage to other property types) AND did not involve spike strips or a tactical raid.
Vehicle damage from spike strips: (1) claim involved only vehicle property damage; AND (2) claim involved spike strips.
Commercial property damage: (1) claim involved only commercial property damage; AND (2) entity filing the claim was a business, lawyer, or insurance company.

21 As noted in the Methods, we excluded from our study car accidents where law enforcement officers were in traffic like any other motorist.

22 This is consistent with what the *Austin American-Statesman* found in its analysis of claims for building damage caused by law enforcement in Austin, Texas. It reported a median value for residential claims of $1,071. Grumet, 2026a.

23 Residential damage from tactical raids accounted for 5 of the 10 largest claims in our dataset.

24 The family also claimed that during the operation, officers evacuated family members from the home and forced them to leave their dog behind. Despite officers' promises to look after the dog, when the family returned home two days later, they found their pet dead without food or water.

25 Based on our human-coded sample, a large percentage of businesses were holding companies, often limited liability corporations, used by small landlords.

26 For approximately 20% of records, it was unclear whether the claimant was the target. We excluded those records from this calculation.

27 The claimant's relationship with the target of the law enforcement action was unclear for 9% of records. We excluded those records from this calculation.

28 For approximately 16% of records, the claim outcome was unclear. In some cases, this was because the claim was still open at the time of the public records request response. In others, the record simply did not provide enough information to ascertain the outcome.

29 In only 2% of records did the government say it was denying the claim because the claimant was the cause of the law enforcement action. However, the F1 score for this response was only 0.5. These percentages do not include roughly 300 claims for which jurisdictions provided determination information in separate spreadsheets, not as part of the claim record itself. Those spreadsheets included the claim outcome and paid amount (if applicable), which we incorporated into our analyses. However, they generally did not include the rationales for the decisions. Also not included are another 14 claims that did not have a predicted denial reason. For further details about our process for incorporating separately provided determination information, see Appendix A.

30 To test whether the claim amount was related to the likelihood of payment, we ran a logistic regression with standard errors clustered by jurisdiction. The log of the claim amount was our independent variable as (1) we suspected the percentage change in the claim amount was more important than the absolute change and (2) model evaluation criteria (Akaike Information Criterion and Bayesian Information Criterion) indicated the logged amount was a better fit. As controls, we included whether the incident involved a tactical raid, pursuit, spike strips, or warrant; the type of claimant; whether the claimant was the target of the law enforcement action; the claimant's relationship with the target; the jurisdiction's region of the country; a log of the jurisdiction's general expenditures; and the year of the incident. We found the logged claim amount was statistically significant with a p-value under 0.001, although the magnitude of the relationship was relatively small. Other models with alternative specifications (e.g., measuring year differently) showed substantively similar results. This analysis is, of course, correlational and subject to limitations common to such designs.

31 This analysis includes only jurisdictions with at least six decided claims for residential damage resulting from a tactical raid where the claimant was not the target.

32 This analysis includes only jurisdictions with at least six decided claims for residential damage resulting from a pursuit where the claimant was not the target.

33 2,752 claims / (222 jurisdictions * 9 years) = 1.38 claims per jurisdiction, per year. This includes the 178 jurisdictions that provided complete claim documentation and the 44 jurisdictions that indicated they had no responsive records.

34 As noted in the Methods, we excluded from our dataset car accidents where law enforcement officers were in traffic like any other motorist, as well as some other kinds of claims not relevant to our study. The codebook in Appendix C lists the kinds of claims we excluded as irrelevant. To derive our $6,476 estimate, we imputed the claim amounts for the roughly 19% of claims for which this information was missing. Specifically, if there was no claim amount (e.g., because the claimant was initially unsure of the cost of damages) but there was a paid amount, we used the paid amount as the claim amount. If there was no paid amount either, we used the average claim amount for the broad claim type. For example, the average claim amount for residential damage that did not involve a pursuit or raid was $2,397. Therefore, if a general residential damage claim had neither a claim amount nor a paid amount, we imputed $2,397.

35 Only Milwaukee and the District of Columbia had higher average claim amounts. Milwaukee's $116,000 average is driven, to a large extent, by a claim for about $550,000; the city had only one other claim over $20,000. The District's $143,000 average owes partly to two large claims totaling $376,000; none of its other claims were over $30,000. Without those outlier claims, Milwaukee's and the District's averages would be about $55,000 and $102,000 per year.

Notably, while Federal Way had only 23 claims total over the nine years, Milwaukee and the District had 194 and 229, respectively, including many expensive vehicle claims.

36 *Sheetz* v. *County of El Dorado*, 601 U.S. 267, 273–74 (2024) (observing that "the Takings Clause saves individual property owners from bearing 'public burdens which, in all fairness and justice, should be borne by the public as a whole.'" (quoting *Armstrong* v. *United States*, 364 U.S. 40, 49 (1960)).

37 Using the U.S. Department of Agriculture's Monthly Cost of Food Reports for low- and moderate-cost food plans for June 2026, we estimate the monthly grocery costs for a family of four range from around $950 to $1,400. See https://www.fna.usda.gov/research/cnpp/usda-food-plans/cost-food-monthly-reports. Median gross monthly housing costs for renters come to $1,487. This figure is from the 2024 U.S. Census Bureau's American Community Survey, Table B25064. See https://www.census.gov/programs-surveys/acs.html and https://data.census.gov/table/ACSDT1Y2024.B25064

38 Bennett, 2026.

39 Complaint at ¶¶ 54, 56, *Hadley* v. *City of South Bend*, No. 71C01-2312-MI-000867 (Ind. Cir. Ct. Dec. 15, 2023), https://ij.org/wp-content/uploads/2023/12/COMPLAINT-Hadley-v.-City-of-South-Bend-et-al.pdf; *Hadley* Petition, *supra* note 11, at 5–6.

40 *Pena* Petition, *supra* note 11, at 2.

41 Institute for Justice. (2023, July 19). *Innocent small business owner seeks compensation after shop destroyed by SWAT team*. https://ij.org/case/los-angeles-swat-destruction/

42 Grumet, 2026a.

43 Grumet, 2026a.

44 Grumet, B. (2026c, June 1). After Austin police leave, vulnerable residents shouldn't be less secure. *Austin American-Statesman*. https://www.statesman.com/opinion/damaged-for-good/article/police-welfare-checks-domestic-opinion-22160555.php

45 Grumet, B. (2026b, May 14). Austin has a process for police damage claims: It always says no. *Austin America-Statesman*. https://www.statesman.com/opinion/damaged-for-good/article/austin-property-damage-claims-denied-22160437.php; Grumet, 2026c.

46 To calculate this percentage, we used the U.S. Census Bureau's 2022 Census of Governments State and Local Government Finance Tables, calculated the total amount of direct general expenditures for each of the 222 responsive local governments in our sample, summed those values, and divided by 222 to arrive at an annual average of $1.25 billion. We then divided the $6,476 average annual claims by the average amount of direct general expenditures, which equaled 0.00052%. Alternate formulations of this analysis—such as using the median amount of direct general expenditures or calculating the percentage on a per jurisdiction basis—still result in percentages of less than one hundredth of a percent. See https://www.census.gov/data/datasets/2022/econ/local/public-use-datasets.html

47 See Schwartz, J. C. (2016). How governments pay: Lawsuits, budgets, and police reform. *UCLA Law Review, 63*(5), 1144–1298. https://www.uclalawreview.org/how-governments-pay-lawsuits-budgets-and-police-reform/

48 We calculated this percentage using the 2022 Census of Governments State and Local Government Finance Tables.

49 These figures exclude claims where the outcome was not present in the record.

50 Minn. Stat. § 626.74.

51 H.F. 4133, 94th Leg., Reg. Sess. (Minn. 2026), https://www.revisor.mn.gov/bills/94/2026/0/HF/4133/?body=house

52 Broadnax, T. C. (2026, June 2). *Payment of no-fault property claims* (Memorandum). City of Austin, Texas. https://services.austintexas.gov/edims/document.cfm?id=474915. See also *American-Statesman* Editorial Board. (2026, June 3). New policy gets police damage claims right. Now comes the hard part. https://www.statesman.com/opinion/editorials/article/police-property-damage-editorial-22287773.php

53 *Pena* v. *City of Los Angeles*, 158 F.4th 1033, 1047 (9th Cir. 2025) (emphasis in original).

54 Brief of Chief Thomas J. Tiderington as Amicus Curiae in Support of Petitioner at 5–6, *Pena* v. *City of Los Angeles*, No. 25-1163 (U.S. May 11, 2026), https://www.supremecourt.gov/DocketPDF/25/25-1163/408545/20260511134431882_25-1163%20Amicus%20Brief.pdf [hereinafter Tiderington Amicus]; Brief of

Chief Thomas J. Tiderington as Amicus Curiae in Support of Petitioner at 5–6, *Hadley* v. *City of South Bend*, No. 25-1158 (U.S. May 8, 2026), https://www.supremecourt.gov/DocketPDF/25/25-1158/408457/20260508153914346_25-1158%20Amicus%20Brief.pdf

55 Tiderington Amicus, *supra* note 54, at 3–4.

56 We found these 243 examples by searching the incident descriptions generated by the AI coder. As a check, we selected 20% of these claims and verified using the original claim documentation that the claimant was referred to a claim form or was told the local government would pay for the damage. All 20% indeed included such statements.

57 Grumet, 2026a.

58 Grumet, 2026a, 2026c.

59 Transcript of Trial Proceedings at 242:8–22, 247:2–248:1, *Baker* v. *City of McKinney*, No. 4:21-cv-00176-ALM (E.D. Tex. June 21, 2022).

60 Grumet, 2026b.

61 *Pumpelly* v. *Green Bay & Miss. Canal Co.*, 80 U.S. (13 Wall.) 166 (1872).

62 *Horne* v. *Dep't of Agric.*, 576 U.S. 351 (2015).

63 *Cedar Point Nursery* v. *Hassid*, 594 U.S. 139 (2021).

64 *Loretto* v. *Teleprompter Manhattan CATV Corp.*, 458 U.S. 419 (1982).

65 Tiderington Amicus, *supra* note 54, at 6 (emphasis in original).

66 Two fields—(1) the damage was caused by law enforcement responding to the wrong address and (2) a person or animal died during the incident—were rare and prone to false positive predictions from the AI coder. A third field—whether a warrant was associated with the incident—was not reliable as the underlying records frequently did not address the presence or lack of a warrant.

67 Occasionally at this stage, we would add administrative language clarifying the format of the output (e.g., "Answer with the date in YYYY-MM-DD format or leave blank, no other text").

68 The two exceptions were our questions about (1) whether the claimant was the target and (2) the claimant's relationship to the target. We submitted these questions to the model as one combined prompt; the model then returned an answer to both questions, with the two answers separated by a comma (e.g., "A,C"). We did this because the answer to the second question was highly dependent on the answer to the first. For example, if the claimant was, in fact, the target of the police action, then the follow-up question regarding the claimant's relationship to the target was not applicable. Similarly, if it was unknown whether the claimant was the target, then the claimant's relationship to the target would necessarily also be unknown.

69 This step was necessary because there is a random component to AI predictions that means that the same question submitted to the same model multiple times will sometimes result in different answers. This is especially true for close calls or edge cases where different pieces of information are pulling the model in different directions.

70 The reflection model was Google's Gemini 3 Pro, with a temperature of 1 and a thinking budget of 15,000.

71 To be Pareto efficient, a candidate prompt could not be "dominated" by another candidate prompt. A prompt dominated another if it improved performance without introducing any new errors. For example, if candidate prompt A correctly predicted records #1–8 but not records #9 and #10, and candidate prompt B predicted all 10 records correctly, then prompt A would be dominated by prompt B because prompt B improved performance without introducing new errors. But if prompt B incorrectly predicted record #3 while correctly predicting the other nine records, then prompt A would *not* be dominated because prompt B introduced a new error. Practically, the use of Pareto efficiency encouraged diversity in the pool of candidate prompts, leading to a wider range of tested strategies.

72 The leading candidates were sampled probabilistically, with the best performing ones having the highest probability of being selected as the starting prompt.

73 The process outlined here generally follows the Genetic-Pareto method in this paper, adapted slightly to fit our study's unique needs and goals. Agrawal, L. A., Tan, S., Soylu, D., Ziems, N., Khare, R., Opsahl-Ong, K., Singhvi, A., Shandilya, H., Ryan, M. J., Jiang, M., Potts, C., Sen, K., Dimakis, A. G., Stoica, I., Klein, D., Zaharia, M., & Khattab, O. (2026). GEPA: Reflective prompt evolution can outperform reinforcement learning. *arXiv*. https://doi.org/10.48550/arXiv.2507.19457

74 The reflection model never explicitly stated these were the strategies it used, of course. Rather, this is a summary of the types of changes the reflection model tended to make.

75 These four criteria were (1) the existence of a claim; (2) the claimant alleged property damage; (3) the property damage was caused by law enforcement; and (4) law enforcement was performing a core law enforcement activity when the damage occurred (e.g., making a traffic stop, driving to a call, executing a warrant, pursuing a suspect, or making an arrest).

76 The relevance subtypes were records that were irrelevant because there was no evidence that a claim was filed or because the claimant was not the owner of the damaged property.

77 The 41 additional records were added to the training data for only the one question where they represented a rare event. For example, a claim added to our training data because it contained a spike strip incident would be added to the training data for only the spike strip question. The remaining fields were human coded but not used for training.

78 The final model was Google's Gemini 2.5 Pro, with a temperature of 1 and a thinking budget of 10,000. We also continued to submit each question for each record as a standalone query and use the majority prediction from three model runs as the final response.

79 We used a four-person committee to mitigate the risk that something would be overlooked. This was necessary because the AI coder was exceptionally good at finding information buried in a record that a human might have missed.

80 The 490 human-coded records comprised the 86 training records, the 86 validation records, the 200 test records, the 41 records used for focused training, two records that generated errors when run through the AI model, and 75 records that were manually coded early in the process for use in initial calibrations. Not included in the human-coded total are the 150 records we added as additional training for relevance, the claim amount, and the determination date. As described above, we manually coded those 150 records for those three questions only. The AI tool coded them for all the other questions.

81 We initially flagged 80 potential duplicates with identical jurisdictions, incident dates, and claim amounts. We manually reviewed those records, resulting in the removal of 11 confirmed duplicates.

82 We used the human-coded records to establish minimum and maximum outlier thresholds. We calculated the thresholds by taking the interquartile ranges of the log-scaled values (to account for the right-skewed nature of the dollar amounts) and then multiplying by 1.5. For the claim amount, the thresholds were below $34 and above $45,065; for the paid amount, the thresholds were below $37 and above $29,037.

About the Authors

DAVID WARREN, PH.D.

David Warren is a former senior research analyst on IJ's strategic research team, where he developed and conducted research on civil forfeiture, the open fields doctrine, takings, and other issues central to IJ's mission. In previous roles at the Brookings Institution and Indiana University, Warren published research in a wide variety of areas, including energy economics, city and regional policy, and federalism. He also previously taught Urban Problems and Solutions at Indiana University, where he introduced undergraduates to the nuts and bolts of local government, including concepts and research related to zoning, policing, education, and other issues that relate to IJ's work. Warren earned a bachelor's degree from the University of Wisconsin-Green Bay, a master's from The Ohio State University, and a Ph.D. in public affairs from Indiana University.

JASON TIEZZI

Jason Tiezzi is a data scientist who frequently collaborates with IJ, including on projects such as *Unaccountable*, *Too Many Licenses?*, the third edition of *License to Work*, and the third and fourth editions of *Policing for Profit*. He holds a bachelor's degree from the College of William and Mary and a master's degree in data science from the University of Virginia.

MINDY MENJOU

Mindy Menjou is assistant director of IJ's strategic research team. She co-authored the fourth edition of *Policing for Profit*, *Beauty School Debt and Drop-Outs*, *Municipal Fines and Fees*, and *Forfeiture Transparency and Accountability*. Menjou holds a bachelor's degree from the University of Southern California and a master's degree from the Humboldt University of Berlin in Germany.

Acknowledgments

The authors are deeply grateful to everyone who made this report possible. Lisa Knepper has served as a guiding light throughout, using her vision and keen eye for detail to greatly improve the final product. Similarly, Dick Carpenter's statistical knowledge and sage wisdom have been pivotal to giving this report its shape.

Chenyu Li was instrumental in executing the AI portion of this project, turning this from a pipe dream to a reality. We thank him for his expertise, thoughtfulness, and patience with our many, many, many questions.

We are also grateful to attorneys Marie Miller, Jeffrey Redfern, and Suranjan Sen, who generously shared their considerable legal knowledge with us and helped us zero in on our key findings.

Allan Hegedus and Rose Quinlan were the backbone of our team, performing the tedious and exacting coding of thousands of pages of records with great accuracy. Allan also helped organize some of our records, performed data checking, and assisted with several key analyses. His contributions and general good humor were greatly appreciated.

Our draft benefited from the thoughtful review of Scott Bullock, Dana Berliner, Robert McNamara, Marie Miller, Jeffrey Redfern, and Suranjan Sen.

Sarah Eckhardt was instrumental in the early days of the project as we figured out how, exactly, we should code claim records that were often lacking in completeness and clarity. Harrison Weeks and Tony Laudadio contributed to early coding efforts, while Elyse Pohl provided feedback on codebook questions.

Obtaining all of these claim records also required a team of diligent colleagues sending FOIA requests and corresponding with records custodians from hundreds of government agencies. That team included Tony Laudadio, Cara Di Silvio, Megan Passon, Sarah Eckhardt, Hannah So, Alec Mena, and Nikhil Agarwal. Christian Stewart helped us convert large files often containing dozens or even hundreds of records into individual sequential records for coding.

Sam Volokh and Isabella Cerqueira added key polish by helping us find excellent examples and locating important context and background information. And Matthew West lent his wealth of expertise by providing feedback on the included regression analysis.

Our superb cover and the color and font schemes were designed by TJ Skindzier. Rima Gerhard created the simple yet effective landing page for the report. And Evan Lisull checked and formatted all the legal citations and proofread the report in full.

About IJ

Founded in 1991, the Institute for Justice is a nonprofit, public interest law firm. Our goal is to end widespread abuses of government power and secure the constitutional rights that allow all Americans to pursue their dreams.

Institute for Justice
901 N. Glebe Road, Suite 900
Arlington, VA 22203

INSTITUTE for JUSTICE

www.ij.org
P: 703.682.9320
F: 703.682.9321