

Certify to Protect? Part 139 Expansion and the Severity Margin in Wildlife Strikes

APEP Autonomous Research* @olafdrw

April 1, 2026

Abstract

Management-based safety regulation may change how organizations manage risk even when it does not reduce incident counts. I study the 2004–2007 expansion of FAA Part 139 certification to newly regulated Class III commuter airports, which imposed wildlife-hazard planning, inspections, and training requirements on airports serving scheduled aircraft with 10–30 seats. Using FAA wildlife-strike records for 1990–2024 and a difference-in-differences design that compares newly certificated airports to never-certificated airports, I find no clear effect on total reported strikes or on the damage share of reported strikes. The most consistent signal is on the severe tail: the preferred Poisson estimate implies about a 72 percent decline in substantial-or-destroyed strikes, although the estimate is imprecise and the treated sample narrows to 20 airports in the matched panel. The evidence is therefore consistent with certification improving severity management more than strike incidence.

JEL Codes: L51, L93, Q58

Keywords: aviation safety, wildlife strikes, management-based regulation, airport certification, difference-in-differences

*Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch (cumulative: 37m).

1. Introduction

Wildlife strikes are a chronic aviation safety problem. Most reported strikes do not damage aircraft, but a small subset causes substantial airframe damage, engine damage, diversions, aborted takeoffs, or destroyed aircraft. That tail risk matters disproportionately for welfare: airports and carriers can tolerate some nuisance-level incidents, but a severe bird or mammal strike can impose large repair costs and operational disruption. The policy question is therefore not only whether regulation reduces strike incidence, but whether it reduces strike severity.

This paper studies a management-based safety reform in U.S. aviation. In February 2004 the Federal Aviation Administration (FAA) revised 14 CFR Part 139 and extended airport-certification requirements to a new group of Class III commuter airports serving scheduled aircraft with 10–30 passenger seats ([Federal Aviation Administration, 2004, 2026](#)). Those airports became subject to the broader Part 139 compliance regime, including wildlife-hazard assessments and management plans when triggering conditions were met, self-inspection requirements, training, and additional operational oversight. The reform is a useful test case for management-based regulation in the sense of [Coglianese and Lazer \(2003\)](#): instead of mandating one narrow technology, it required airports to build internal safety-management routines around an identified risk.

The main empirical challenge is that certification can affect reporting as well as underlying risk. If newly certificated airports become more diligent about documenting strikes, total reported strikes may rise even when the true hazard environment improves. A design focused only on incidence could therefore misread the policy. I address that problem by separating four outcomes: total reported strikes, damaging strikes, severe strikes, and the damage share among reported strikes. The distinction is central rather than auxiliary.

I assemble a panel from the FAA wildlife-strike public API and FAA airport metadata for 1990–2024. Treatment is defined using the FAA’s own regulatory-evaluation appendix listing the airports newly brought into Class III certification by the 2004 rule ([Federal Aviation Administration, 2004](#)). Because compliance ran through June 2007, I treat 2007 onward as post and omit the 2004–2006 transition years. The baseline specification is an airport-year difference-in-differences with airport and year fixed effects, Poisson estimators for sparse counts, and airport-clustered standard errors.

The evidence is narrower than the original idea manifest suggested, but still informative. After matching the historical Class III roster into the strike panel, the treated sample falls to 20 airports, exactly clearing the minimum sample floor for a V1 paper. This is therefore not a full-universe Class III study; it is a matched historical-roster exercise. In that matched

sample, total reported strikes do not move clearly after certification. The baseline Poisson estimate for total strikes is -0.164 (s.e. 0.696), and the estimate for damaging strikes is -0.306 (s.e. 0.360). Damage-share regressions are essentially zero. The only outcome that shows a stable negative signal is severe strikes: the baseline estimate is -1.275 log points (s.e. 0.756), which corresponds to roughly a 72 percent decline in substantial-or-destroyed strikes. Raw means point in the same direction, but the event count is small. Severe strikes fall from 13 events in 280 treated airport-years before 2004 to 2 events in 360 treated airport-years after 2007, while the control group falls from 356 to 196 events over much larger exposure.

I do not interpret that result as proof that Part 139 broadly reduced strike incidence. The design is more convincing on the severity margin than on overall strike counts. Both treated and control airports report more strikes over time, consistent with rising reporting intensity in the national database ([Federal Aviation Administration, 2024](#)). The contribution of the paper is therefore disciplined downshifting rather than overclaiming: the evidence is consistent with newly imposed wildlife-management obligations reducing the upper tail of strike severity, but not with a clean reduction in total strike incidence.

The paper contributes to three literatures. First, it adds to work on management-based regulation by showing that internal compliance systems may matter most on high-consequence outcomes rather than on overall event frequency ([Coglianese and Lazer, 2003](#)). Second, it contributes to the aviation-wildlife literature, which has documented the scale of strike risk and its operational consequences but has had less to say about causal effects of airport regulation ([Burger, 1985](#); [Washburn et al., 2013](#); [Altringer et al., 2022, 2023](#)). Third, it applies a transparent difference-in-differences design with a fixed treatment date chosen to respect the policy’s implementation window, following modern concerns about treatment timing and dynamic comparisons ([Callaway and Sant’Anna, 2021](#); [Sun and Abraham, 2021](#); [Goodman-Bacon, 2021](#)).

The rest of the paper proceeds as follows. Section 2 describes the institutional setting. Section 3 presents the data and sample construction. Section 4 lays out the empirical design. Section 5 reports the main results and robustness checks. Section 6 discusses interpretation and limitations. Section 7 concludes.

2. Institutional Background and Policy Setting

2.1 Part 139 and Why Certification Could Matter

Part 139 governs certification requirements for U.S. airports serving certain scheduled air-carrier operations. Certification is not a wildlife-specific rule. It is a broader safety regime covering inspection routines, emergency planning, airfield conditions, personnel qualifications,

and wildlife-hazard management when risks trigger the relevant provisions. The mechanism in this paper is therefore organizational. Certification can force airports to formalize routines that were previously ad hoc: inspect more systematically, document hazards, train staff, contract for mitigation, and update plans after damaging events.

That feature makes the reform an attractive case for management-based regulation. The regulator does not perfectly observe local wildlife risk at each airport and cannot write one efficient technical rule for all airports. Instead, it compels the airport operator to build internal systems that identify hazards and respond to them (Coglianese and Lazer, 2003). If that logic works, the most likely effect is not necessarily fewer animal-aircraft interactions overall. Wildlife presence depends on local ecology, land use, migration paths, and traffic volume. The more plausible channel is fewer high-severity events conditional on operating in a risky environment.

2.2 The 2004 Expansion to Class III Airports

The relevant policy shock is the FAA’s 2004 final rule revising Part 139. The rule expanded certification to airports serving scheduled operations by aircraft with 10–30 passenger seats, the new Class III category, and the compliance period ran into mid-2007 (Federal Aviation Administration, 2004). The FAA’s regulatory evaluation explicitly lists the airports expected to become newly certificated under the rule, which provides a historically grounded treatment roster rather than an inferred one.

This historical roster matters because current certification-status spreadsheets are not a clean source of treatment assignment. Some airports changed status later, some airports closed, and current class labels do not reveal whether an airport was part of the original 2004 expansion. I therefore treat the regulatory-evaluation appendix as the primary source for treatment assignment and use current FAA metadata only for cross-checking airport codes and locations.

2.3 Why Reporting Could Confound the Design

The reporting problem is first-order in the wildlife-strike setting. The FAA database reflects both realized strikes and airports’ willingness or capacity to document them. Certification may improve compliance culture and reporting discipline even if underlying wildlife risk is unchanged. For that reason, a rise in total strike reports after certification would not necessarily imply policy failure, and a null total-strike estimate would not necessarily imply no safety benefit.

That logic motivates the outcome decomposition used below. If certification only increases

reporting, total reports should rise without a corresponding drop in severity. If certification improves on-airport wildlife management, one should see the largest movement in damaging or severe strikes, especially at the upper tail where mitigation matters most.

3. Data

3.1 Data Sources

The main outcome data come from the FAA National Wildlife Strike Database public interface, which exposes a searchable API for strike-level records ([Federal Aviation Administration, 2024](#)). I query the public database for 1990–2024 and merge the resulting strike records to FAA airport metadata and certification information. The unit of analysis is the airport-year.

Treatment assignment comes from the FAA’s 2004 regulatory-evaluation appendix, which lists 37 airports expected to enter the new Class III category ([Federal Aviation Administration, 2004](#)). I manually transcribed that roster into a local file and then merged it into the FAA airport-code universe. Not all historically listed airports survive into the usable matched panel. Some have sparse or unmatched strike data, and at least one closed before the policy’s effective post period. The final matched panel contains 20 treated airports and 757 control airports.

Controls are U.S. airports that are never treated, have current Part 139 class 0 or blank in the FAA metadata, and have at least one pre-2004 strike observation. That restriction is designed to avoid comparing newly certificated airports to locations with no measurable baseline wildlife-strike reporting at all. The analysis window uses 1990–2003 as the pre period, drops 2004–2006 as transition years, and treats 2007–2024 as post.

3.2 Variables

I construct four primary outcomes. *Total strikes* counts all strike reports in an airport-year. *Damaging strikes* counts strikes with recorded aircraft damage codes corresponding to minor, uncertain minor, substantial, or destroyed damage. *Severe strikes* counts only substantial or destroyed damage. *Damage share* is the share of reported strikes in an airport-year that involve any recorded damage. The severe-strike outcome is the cleanest measure of the policy’s upper-tail safety consequences, while the damage share is the most direct attempt to net out pure reporting expansion.

The main treatment variable equals one for historically listed Class III airports in years 2007 onward. Airport and year fixed effects absorb time-invariant airport heterogeneity and national shocks to traffic, reporting, and wildlife conditions.

3.3 Summary Statistics

Table 1: Sample composition and pre-period strike rates

| Group | Airports | Pre total strikes | Pre damaging strikes | Pre damage share |
|------------------|----------|-------------------|----------------------|------------------|
| Treated airports | 20 | 0.411 | 0.107 | 0.082 |
| Control airports | 757 | 0.304 | 0.112 | 0.078 |

Table 1 shows that treated airports were somewhat more strike-prone than controls even before certification. Treated airports averaged 0.411 total strikes and 0.107 damaging strikes per year in the pre period, compared with 0.304 and 0.112 for controls. The pre-period damage share is similar across groups: 0.082 at treated airports and 0.078 at controls. The treated group is therefore not a low-risk set of commuter airports that happened to receive certification. If anything, the treated airports enter the policy with slightly higher total strike incidence.

The raw post-period means already hint at the paper’s central decomposition. Total reported strikes rise in both groups, from 0.411 to 0.825 among treated airports and from 0.304 to 0.719 among controls. Damaging and severe strikes, however, fall in both groups and fall more sharply in the treated sample. Because these are unconditional means, they are descriptive rather than causal. Still, they motivate focusing on whether certification changed severity more than incidence.

4. Empirical Strategy

4.1 Baseline Difference-in-Differences Design

The baseline design is a two-group, two-period difference-in-differences embedded in a long airport-year panel:

$$Y_{at} = \beta(\text{Treated}_a \times \text{Post}_t) + \alpha_a + \gamma_t + \varepsilon_{at}, \quad (1)$$

where a indexes airports and t indexes years. α_a and γ_t denote airport and year fixed effects. The coefficient of interest, β , measures the post-2007 change at historically newly certificated airports relative to never-certificated airports.

I estimate count outcomes with Poisson fixed effects because the outcomes are sparse, nonnegative, and include many zeros. Damage share is estimated with linear fixed effects weighted by the number of reported strikes in the airport-year. Standard errors are clustered at the airport level.

The identifying assumption is that, absent certification, newly certificated and never-certificated airports would have experienced parallel changes in strike severity and incidence after 2007. Because the treatment date is effectively common once the transition window is omitted, the design avoids the negative-weight pathologies that arise in staggered-adoption two-way fixed effects settings (Goodman-Bacon, 2021). I nevertheless cite the modern DiD literature because it clarifies why dynamic comparisons are difficult in short or contaminated transition windows (Callaway and Sant’Anna, 2021; Sun and Abraham, 2021).

4.2 Design Choices and Threats

Three choices are doing important work. First, I set the post period to begin in 2007 rather than 2004 because the rule’s compliance window extended through June 2007 (Federal Aviation Administration, 2004). Treating 2004 as an immediate post year would contaminate the estimates with partial compliance. Second, I rely on the historical Class III appendix instead of today’s certification roster because the current spreadsheet is not a treatment roster. Third, I center the design on severity outcomes to address the reporting objection directly.

The design still has clear limitations. The treated sample is modest, with only 20 matched airports. Airport traffic, fleet mix, surrounding land use, and local wildlife ecology may evolve differentially over time. The control group is large but imperfect: never-certificated airports are not ideal counterfactuals for airports serving scheduled 10–30 seat aircraft, and I do not observe a common operations denominator for the full panel. Finally, the FAA database itself has become more comprehensive over time, which is why null incidence results do not by themselves settle the policy question.

5. Results

5.1 Main Estimates

Table 2 presents the main estimates. Column 1 shows no clear reduction in total reported strikes. The point estimate is -0.164 log points with a standard error of 0.696, far too imprecise to distinguish between moderate declines and moderate increases. This is exactly the sort of result one would expect if certification changed reporting, if underlying strike risk remained noisy, or both.

Column 2 turns to damaging strikes. The point estimate is -0.306 log points (s.e. 0.360), which corresponds to roughly a 26 percent decline, but again the estimate is imprecise. Column 3 isolates severe strikes and produces the sharpest result in the paper: -1.275 log

Table 2: Main results

| Dependent Variables: | total_strikes | damaging_strikes | severe_strikes | damage_share |
|-----------------------|-------------------|-------------------|-------------------|--------------------|
| | Total strikes | Damaging strikes | Severe strikes | Damage share |
| Model: | (1) | (2) | (3) | (4) |
| | Poisson | Poisson | Poisson | OLS |
| <i>Variables</i> | | | | |
| treated_postTRUE | -0.164 (0.696) | -0.306 (0.360) | -1.27* (0.756) | -0.0005 (0.015) |
| <i>Fixed-effects</i> | | | | |
| airport_fe | Yes | Yes | Yes | Yes |
| year_fe | Yes | Yes | Yes | Yes |
| <i>Fit statistics</i> | | | | |
| Observations | 24,896 | 22,016 | 10,788 | 24,896 |
| R ² | | | | 0.09326 |

Clustered (airport_fe) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

points (s.e. 0.756), equivalent to roughly a 72 percent decline. Given the modest treated sample, I treat that magnitude as suggestive rather than definitive. It is based on very few treated severe events: 13 in the pre period and 2 in the post period. The estimate is only marginally significant and should be read together with the robustness table rather than in isolation.

Column 4 asks whether certification changed the composition of reported strikes by looking at the damage share. The answer is basically no. The coefficient is -0.0005 (s.e. 0.015), close to mechanical zero. That null is important. It implies the evidence does not support a clean story in which certification simply lowered the damage probability of the average reported strike. One reason the share may remain flat even when severe strikes decline is arithmetic: the severe category is tiny relative to all reported strikes, and reporting growth expands the denominator. The upper-tail result is therefore stronger than the average-intensity result.

5.2 Outcome Decomposition

Table 3 makes the reporting-versus-severity logic more transparent. Total reports rise strongly over time in both groups. That is consistent with broader increases in strike reporting documented by the FAA and in related work ([Federal Aviation Administration, 2024](#); [Altringer et al., 2023](#)). Damaging strikes move in the opposite direction: treated

Table 3: Outcome decomposition by group and period

| Group-period | Mean total strikes | Mean damaging strikes | Mean damage share |
|--------------------|--------------------|-----------------------|-------------------|
| Control, pre-2004 | 0.304 | 0.112 | 0.078 |
| Control, post-2007 | 0.719 | 0.087 | 0.045 |
| Treated, pre-2004 | 0.411 | 0.107 | 0.082 |
| Treated, post-2007 | 0.825 | 0.061 | 0.032 |

airports decline from 0.107 to 0.061 damaging strikes per year, while controls decline from 0.112 to 0.087. Severe strikes, not shown directly in the decomposition table, fall from 0.046 to 0.006 among treated airports and from 0.034 to 0.014 among controls.

The decomposition therefore supports a bounded interpretation. Certification did not produce an obvious break in the number of reported strike events. Instead, the strongest movement occurs in the most consequential events. That pattern is consistent with a safety-management mechanism that improves hazard response and damage mitigation more than it changes the underlying exposure of airports to birds and mammals.

5.3 Robustness

Table 4: Robustness checks

| Dependent Variables: | severe_strikes | damaging_strikes | severe_strikes | damage_share |
|-----------------------|--------------------------------|--------------------------------|----------------------------|-----------------------------|
| Model: | Sev., restr. (1) Poisson | Dmg., restr. (2) Poisson | Sev. + trend (3) OLS | Share + trend (4) OLS |
| <i>Variables</i> | | | | |
| treated_postTRUE | -1.28* (0.756) | -0.312 (0.360) | -0.020 (0.032) | 0.020 (0.037) |
| <i>Fixed-effects</i> | | | | |
| airport_fe | Yes | Yes | Yes | Yes |
| year_fe | Yes | Yes | Yes | Yes |
| <i>Fit statistics</i> | | | | |
| Observations | 10,757 | 21,984 | 24,896 | 24,896 |
| R ² | | | 0.06410 | 0.09327 |

Clustered airport standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 4 reports four checks targeted at the main design concerns. Column 1 restricts the severe-strike sample using the same trimmed comparison set used elsewhere in the robustness exercise. The severe-strike estimate is essentially unchanged at -1.28 log points (s.e. 0.756). Column 2 applies the same restriction to damaging strikes and produces -0.312 (s.e. 0.360), again very close to the baseline damage estimate.

Columns 3 and 4 add treated-specific linear trends in linear fixed-effects models. The severe-strike coefficient becomes -0.020 (s.e. 0.032), and the damage-share coefficient becomes 0.020 (s.e. 0.037). These trend-adjusted linear models are much less informative for the sparse upper-tail counts than the Poisson specifications, but they are still useful as a discipline device: the severe result is not so mechanically tied to one narrow coding choice that it flips into a large positive estimate under alternative trend controls.

Taken together, the robustness exercises do not transform the severe-strike estimate into a precise theorem. They instead show that the paper’s narrow claim survives reasonable alternative specifications better than any broad claim about strike incidence would.

6. Discussion

The evidence fits a simple interpretation of management-based regulation. Part 139 certification likely changed what airports did after risk was identified: more formal inspections, staff procedures, wildlife-hazard planning, and mitigation responses. Those changes need not reduce the total number of bird or mammal encounters. Wildlife activity around airports is shaped by geography, seasonality, habitat, migration, and traffic patterns that airport managers only partly control ([Burger, 1985](#); [Washburn et al., 2013](#)). But management routines can still matter on the extensive consequences of those encounters, especially by reducing the probability that a strike escalates into a severe damage event.

At the same time, the paper does not justify a triumphalist reading of the rule. The treated sample is small, the control group is imperfect, and the damage-share estimates are null. The right conclusion is not that Part 139 “solved” wildlife strikes at commuter airports. It is that, in a thin historical matched sample, the available evidence is more compatible with a severity-margin effect than with an incidence effect. That is a useful distinction for aviation policy. If regulators expect wildlife-management rules to eliminate strike counts, they may judge the policy too harshly. If the goal is instead to reduce catastrophic outcomes, the rule may look more successful.

This distinction also matters for evaluation design in other management-based regulations. When compliance systems alter documentation, inspections, and internal controls, raw event counts are often a poor standalone measure of success. Researchers need outcomes that separate reporting intensity from the tail consequences that regulation is actually meant to prevent.

7. Conclusion

I study the FAA's 2004–2007 expansion of Part 139 certification to newly regulated Class III commuter airports using 1990–2024 wildlife-strike data. The paper finds no clear evidence that certification reduced total reported strikes, and essentially no evidence that it changed the overall damage share of reported strikes. The strongest and most stable signal is a decline in severe strikes.

That pattern is consistent with a management-based regulation that improves how airports manage the consequences of wildlife hazards more than how often wildlife strikes are reported. Because the treated sample narrows to 20 airports, the estimates should be read as bounded evidence rather than a final word. Even so, the results suggest that airport safety regulation may show up first on the upper tail of harm.

Acknowledgements

This paper was autonomously generated using Claude Code as part of the Autonomous Policy Evaluation Project (APEP).

Project Repository: <https://github.com/SocialCatalystLab/ape-papers>

Contributors: @olafdrw

First Contributor: <https://github.com/olafdrw>

References

- Altringer, Levi, Sammy Zahran, Stephanie A. Shwiff, Michael J. Begier, and Aaron Anderson**, “Spillover delay effects of damaging wildlife strike events at U.S. airports,” *Economics of Transportation*, 2022, *30*, 100252.
- , **Sophie C. McKee, Jason D. Kougher, Michael J. Begier, and Stephanie A. Shwiff**, “The impact of the COVID-19 pandemic on wildlife–aircraft collisions at US airports,” *Scientific Reports*, 2023, *13* (1).
- Burger, Joanna**, “Factors affecting bird strikes on aircraft at a coastal airport,” *Biological Conservation*, 1985, *33* (1), 1–28.
- Callaway, Brantly and Pedro H. C. Sant’Anna**, “Difference-in-Differences with multiple time periods,” *Journal of Econometrics*, 2021, *225* (2), 200–230.
- Coglianese, Cary and David Lazer**, “Management-Based Regulation: Prescribing Private Management to Achieve Public Goals,” *Law & Society Review*, 2003, *37* (4), 691–730.
- Federal Aviation Administration**, “Final Regulatory Evaluation, Final Regulatory Flexibility Determination, Trade Impact Assessment, and Unfunded Mandates Assessment: 14 CFR Part 139, Certificated Airports,” FAA regulatory evaluation 2004. Accessed April 1, 2026.
- , “National Wildlife Strike Database,” FAA public database portal 2024. Public database accessed via FAA web API on April 1, 2026.
- , “Part 139 Airport Certification,” FAA webpage 2026. Accessed April 1, 2026.
- Goodman-Bacon, Andrew**, “Difference-in-differences with variation in treatment timing,” *Journal of Econometrics*, 2021, *225* (2), 254–277.
- Sun, Liyang and Sarah Abraham**, “Estimating dynamic treatment effects in event studies with heterogeneous treatment effects,” *Journal of Econometrics*, 2021, *225* (2), 175–199.
- Washburn, Brian E., Paul J. Cisar, and Travis L. DeVault**, “Wildlife strikes to civil helicopters in the US, 1990–2011,” *Transportation Research Part D: Transport and Environment*, 2013, *24*, 83–88.

A. Data Appendix

The FAA wildlife-strike public interface exposes strike-level records through the same backend used by the agency’s searchable web portal ([Federal Aviation Administration, 2024](#)). I queried the public database for 1990–2024 and retained U.S. airport observations that could be matched to FAA airport metadata. Strike records report the airport field, event year, and aircraft-damage coding used to construct the outcome variables. Airport metadata provide multiple code systems, including FAA and ICAO identifiers, which are necessary because strike records are not perfectly standardized.

Treatment assignment required a historically grounded roster rather than a current snapshot. I therefore used the FAA regulatory evaluation accompanying the 2004 Part 139 rule, which lists the airports expected to become newly certificated under the Class III expansion ([Federal Aviation Administration, 2004](#)). I manually transcribed the airport codes from that appendix and merged them to the airport metadata. The historical list contains 37 airports, but only 20 appear in the final matched airport-year panel after restricting to valid U.S. airport identifiers and observed strike histories. The 17 historical roster codes that do not survive into the estimating sample are SOW, ELD, BPK, IPL, IYK, LWT, MLS, SDY, ALM, CNM, GUP, SVC, CNY, VEL, BLF, FAQ, and Z08.

The final panel contains 777 airports observed annually from 1990–2024, with the 2004–2006 transition years omitted in estimation. That leaves 24,896 airport-year observations in the main specifications, although Poisson models on sparse outcomes can drop observations through fixed-effect separation. The severe-strike model therefore has fewer estimating observations than the total-strike model.

B. Identification Appendix

The design is deliberately conservative about timing. The rule was issued in 2004, but the compliance period for newly certificated airports ran through June 2007. Coding treatment as starting immediately in 2004 would build partial compliance into the post period. I therefore use 1990–2003 as pre, omit 2004–2006, and start the post period in 2007.

I constructed event-study models during the analysis stage, but I do not present them in the paper. With only 20 treated airports in the matched panel and a three-year implementation window removed from the middle of the series, dynamic plots are not informative enough to carry interpretive weight. The paper is stronger when it focuses on transparent level estimates and targeted robustness checks.

The main remaining identification threat is differential trend selection between newly

certificated airports and the never-certificated control pool. I partially address that threat through fixed effects, by requiring control airports to have a pre-period strike history, and through restricted-sample robustness checks. These steps improve comparability but do not eliminate all concerns.

C. Robustness Appendix

The robustness table in the main text emphasizes two concerns. First, perhaps the severe-strike estimate is driven by a thin set of unusually active control airports. Restricting the comparison set by pre-period strike intensity leaves the severe estimate essentially unchanged. Second, perhaps treated airports were already drifting differently along a linear trend. Adding treated-specific trends attenuates the severe estimate in linear models, but it does not create a positive counter-narrative.

I do not present a long catalog of alternative estimators because the data do not support that kind of rhetorical expansion. The sample is small and the sharpest outcome is very sparse. The paper is better served by reporting a handful of transparent checks and keeping the claim narrow.

D. Additional Figures and Tables

This V1 paper includes no figures. The main text reports four tables: sample composition, main estimates, outcome decomposition, and robustness checks. Appendix Table 5 provides the mandatory standardized-effect-size summary.

E. Standardized Effect Sizes

Table 5: Standardized effect sizes

| Outcome | $\hat{\beta}$ | SE | SD(Y) | SDE | SE(SDE) | Classification |
|-------------------------------|---------------|-------|-------|--------|---------|-------------------|
| <i>Panel A: Pooled</i> | | | | | | |
| Total strikes | -0.001 | 0.397 | 1.191 | -0.001 | 0.334 | Null |
| Damaging strikes | -0.021 | 0.029 | 0.397 | -0.052 | 0.073 | Moderate negative |
| Severe strikes | -0.022 | 0.015 | 0.191 | -0.113 | 0.076 | Moderate negative |
| Damage share | -0.000 | 0.015 | 0.259 | -0.002 | 0.058 | Null |
| <i>Panel B: Heterogeneous</i> | | | | | | |
| Damaging strikes (restricted) | -0.021 | 0.029 | 0.397 | -0.054 | 0.073 | Moderate negative |
| Damage share with trend | 0.020 | 0.037 | 0.259 | 0.076 | 0.144 | Moderate positive |

Notes: **Country:** United States. **Research question:** Does the 2004-2007 expansion of FAA Part 139 certification to newly regulated Class III commuter airports change wildlife-strike incidence and severity at treated airports relative to never-certificated airports. **Policy mechanism:** The rule required newly certificated airports serving 10–30 seat scheduled service to adopt Part 139 compliance systems, including wildlife hazard management, self-inspections, training, and related safety procedures. The key empirical question is whether those new compliance obligations reduced damaging strikes or instead mostly changed reporting intensity. **Outcome definition:** Airport-year counts of total reported strikes, damaging strikes, severe strikes, and damage shares from the FAA Wildlife Strike Database public search extract. **Treatment:** Binary indicator equal to one for airports listed in the FAA’s historical Class III appendix and years 2007 onward after the compliance transition. **Data:** FAA Wildlife Strike Database public API, 1990–2024, merged to FAA airport metadata; airport-year panel with treated and never-certificated control airports. **Method:** Airport and year fixed-effects models estimated with Poisson or linear estimators; standard errors clustered by airport. **Sample:** U.S. airports observed in the FAA strike database, using the historical Class III roster as treated units and never-certificated airports with pre-period strike history as controls. $SDE = \hat{\beta}/SD(Y)$ where $SD(Y)$ is the pre-treatment standard deviation. Classification refers to magnitude, not statistical significance: Large ($|SDE| > 0.15$), Moderate (0.05–0.15), Small (0.005–0.05), Null (< 0.005).