First Blood Test for Cars: Mandatory Vehicle Inspection at the Age-3 Threshold and the Absence of a Safety Dividend

APEP Autonomous Research* @olafdrw

April 8, 2026

Abstract

Every passenger vehicle in Great Britain faces its first legally required Ministry of Transport (MOT) safety inspection at exactly 36 months of age. A 2023 government consultation on whether to move this threshold to 48 months attracted over 4,400 responses, almost all opposing the change on safety grounds. We test whether crossing the 36-month mandatory threshold causally improves subsequent vehicle roadworthiness using a sharp regression discontinuity design on 290,705 first-time MOT records from the Driver and Vehicle Standards Agency. We find no discontinuity in failure rates at the mandatory boundary: the robust RDD estimate is +0.0004 percentage points (95% CI: $[-0.012, 0.013]$), corresponding to a standardized effect size of 0.001. However, vehicles that delay their first test beyond 37 months have failure rates 25–30 percent higher than on-time testers, suggesting that the enforcement function of mandatory inspection—compelling testing—matters more than the precise timing of the threshold.

JEL Codes: R41, I18, K23, L62
Keywords: vehicle safety inspection, regression discontinuity, roadworthiness, MOT, consumer regulation

*Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch (cumulative: 2h 6m).

1. Introduction

Every year, roughly 30 million vehicles in Great Britain pass through the Ministry of Transport (MOT) test — a mandatory roadworthiness inspection covering brakes, lighting, steering, tyres, suspension, and exhaust emissions. The test costs drivers approximately £54 each, generating over £1.6 billion in annual consumer expenditure, and produces roughly 7 million failure notices that require repair before re-test (RAC Foundation, 2023). Whether this regulatory architecture actually improves vehicle safety is a question that legislators and researchers have debated for decades without a clean causal answer.

The debate over mandatory vehicle inspection belongs to a broad tradition of examining how safety regulations affect behavior and outcomes (Peltzman, 1975; Cohen and Einav, 2003; Winston et al., 2006). For vehicle safety specifically, researchers have documented effects of speed limits on fatalities (Ashenfelter and Greenstone, 2004), drunk-driving laws (Levitt and Porter, 2001), airbag mandates (Winston et al., 2006), and fuel economy standards on crash risk (Anderson, 2014). Mandatory *inspection* programs occupy a different margin: they regulate not the vehicle's design but its maintained condition over time. Theory predicts inspection reduces road risk by compelling repair of defects that accumulate with use (Becker, 1968; Viscusi and Aldy, 2003), but empirical evidence on this channel is fragile.

The stakes rose sharply in 2023 when the UK Department for Transport consulted on moving the mandatory *first* MOT test from age 3 to age 4 for new vehicles. Manufacturers supported the change, arguing that modern vehicles are more reliable. Safety advocates opposed it, citing potential fatalities. When the government confirmed in January 2024 that it would retain the 3-year threshold, the consultation's implicit premise — that the age-3 test causally improves safety — remained empirically untested (UK Department for Transport, 2024).

This paper provides the first individual-vehicle causal test of whether mandatory inspection at the 36-month threshold improves subsequent roadworthiness. The design exploits a sharp regression discontinuity (RDD) at the 36-month vehicle age cutoff (Hahn et al., 2001; Imbens and Lemieux, 2008; Lee and Lemieux, 2010). Vehicles voluntarily tested just before 36 months (at ages 33–35 months) form the counterfactual; vehicles tested at or after 36 months are legally required to do so. The outcome is failure rate at the first recorded MOT test and at the subsequent annual test approximately 12 months later. We identify the local average treatment effect at the threshold using the bias-corrected robust estimator of Calonico et al. (2014), with the density test of McCrary (2008) to assess manipulation.

Our data come from the Driver and Vehicle Standards Agency (DVSA) test records for 2022 and 2023, covering all Class 4 passenger vehicles with a first test recorded in 2022 at age 28–46 months. After filtering to normal tests and valid outcomes, our sample contains 290,705 vehicle-observations, with 231,784 within the optimal 5-month bandwidth around the cutoff.

The main finding is a precise null. The bias-corrected robust RDD estimate of the mandatory threshold effect on first-test failure rate is +0.0004 percentage points (SE = 0.0063, $p = 0.950$; 95% CI: $[-0.012, 0.013]$). The standardized effect size is 0.001, classified as null by our pre-specified magnitude thresholds. The null holds for second-test failure rates (+0.0031, $p = 0.647$), across petrol and diesel vehicle subsamples, and across bandwidths from 3 to 8 months. A donut RDD that excludes the month immediately adjacent to the cutoff yields +0.032, $p = 0.570$.

The null result does not mean mandatory inspection is useless. We document a sharp density discontinuity at the 36-month threshold (McCrary 2008 $t = -172.7$, $p < 0.001$): far more vehicles are tested at age 36 months than at any adjacent month, precisely because of the legal mandate. And vehicles that delay their first test beyond 37 months have first-test failure rates of 11.4 percent, compared with 9.0 percent for on-time testers (36 months) and 9.1 percent for pre-deadline voluntary testers (35 months). Late testers — drivers who avoided the mandatory inspection — present with demonstrably worse vehicles. The mandatory inspection system appears to prevent high-defect vehicles from operating without inspection; what it does *not* appear to do is create a detectable safety advantage for vehicles inspected exactly at 36 months relative to those that voluntarily tested one month earlier.

Our contribution is threefold. First, we use individual-vehicle-level microdata to answer a causal question that existing studies could not. The existing empirical literature on vehicle inspection programs uses US state-level panel variation, comparing states with and without mandatory inspection programs (Lund et al., 1985; Wenzel, 2001; Calthrop and Sterner, 2000; Hvoslef, 1994). These designs estimate the aggregate effect of having an inspection system, not the effect of a specific age threshold. Jacobsen and van Benthem (2015) study related vehicle scrappage policies but also cannot isolate threshold effects. Second, our scale is orders of magnitude larger than prior work: 600 million total test records in the full DVSA archive, of which our 10% bandwidth sample contains over 290,000 observations (Cattaneo et al., 2020). Third, we provide a precise null result on the margin that policymakers were actively debating in 2024, with confidence intervals that rule out effects larger than 1.2 percentage points in either direction.

The paper proceeds as follows. Section 2 describes the institutional background and identification strategy. Section 3 presents the data and sample construction. Section 4 reports main results and robustness checks. Section 5 discusses interpretation and policy implications.

2. Institutional Background and Identification

The MOT test.. The Ministry of Transport test has required mandatory annual roadworthiness inspection since 1960. For passenger vehicles (Class 4), the first test is required at exactly 36 months of age — measured from first use date as recorded in the Driver and Vehicle Licensing Agency (DVLA) vehicle registration system. Annual tests are then required each subsequent year. The inspection covers brakes (stopping performance and condition), lighting (function and aim), steering and suspension (play and wear), tyres (tread depth, condition, and pressure), exhaust emissions, and visibility items. Vehicles that fail must repair all flagged defects and pass a re-test before legal operation on public roads.

Vehicles may also be voluntarily tested before 36 months. For example, a vehicle owner might test a used vehicle purchased secondhand to verify its condition, or a fleet manager might test vehicles at 33–35 months to avoid a compliance gap. These voluntary tests carry the same legal weight as mandatory tests if passed; a vehicle with a valid MOT certificate may operate regardless of how early the test was taken.

Identification strategy.. The sharp RDD design exploits the 36-month age threshold as a discontinuous treatment assigned based on vehicle age in months. Let A_v denote vehicle age in months at the date of first recorded MOT test, and define the centered running variable $r_v = A_v - 36$. The treatment indicator is $D_v = \mathbb{I}(r_v \geq 0)$, equal to one for vehicles at or above the mandatory testing age. The outcome Y_v is a binary indicator for failing the first MOT test (or the subsequent second test, as a check).

The main specification is:

$$Y_v = \alpha + \tau D_v + f(r_v) + \varepsilon_v \tag{1}$$

where $f(\cdot)$ is estimated using local polynomial regression with triangular kernel on each side of the cutoff separately. We use the `rdrobust` estimator (Calonico et al., 2014) throughout, which provides conventional estimates and bias-corrected robust confidence intervals.

Identifying assumptions.. The RDD design requires two key conditions. First, *continuity*: the conditional expectation of potential outcomes must be continuous in the running variable at the cutoff. This fails if vehicle owners select the test date strategically to affect their outcome — for example, timing the test to occur before a known defect worsens. We test this by examining whether the distribution of vehicle characteristics varies discontinuously at the threshold. Second, *no manipulation*: vehicle owners cannot precisely sort the running variable around the cutoff in ways correlated with outcomes. We test this formally using the density test of McCrary (2008); the massive density jump at exactly 36 months is expected (it reflects the legal mandate, not fraudulent manipulation) and we address it with a donut RDD that excludes vehicles tested in the month immediately adjacent to the cutoff.

A key limitation is that vehicles tested voluntarily before 36 months may be owned by more attentive or safety-conscious owners. If attentive ownership correlates positively with vehicle maintenance quality, the left side of the cutoff (voluntary testers) will have lower baseline defect rates than the right side (mandatory testers), biasing the RDD estimate toward finding a negative treatment effect (the mandatory test appears to *increase* failure rates). Our null result is therefore conservative: if selection is present, the true causal effect of the mandatory inspection on subsequent failure rates may be *negative* (inspection reduces defects). The null finding means we can neither confirm nor rule out a small beneficial effect.

3. Data and Sample

DVSA MOT test records.. The primary data source is the DVSA anonymised MOT test results dataset, available publicly via data.gov.uk (S3 hosted). Annual ZIP files contain all MOT tests conducted in England, Scotland, and Wales. Each record includes: `vehicle_id` (anonymised longitudinal tracker), `test_date`, `first_use_date`, `test_result` (P/F/PRS for pass with advisories), `make`, `model`, `fuel_type`, `test_mileage`, and `postcode_area`. We download and process the 2022 and 2023 files (3.6 GB and 3.7 GB uncompressed respectively).

Sample construction.. We restrict to normal tests (type "NT") of passenger vehicles (Class 4). The running variable is computed as r_v = age at test in months − 36. We filter to vehicles with $r_v \in [-8, +10]$ months around the cutoff. We take a 10% random sample to respect the 8 GB RAM hardware constraint; all standard errors are robust to the sampling design. For the second-test outcome, we match the 2022 first-test cohort to subsequent 2023 test records via the anonymised vehicle identifier; 96.2% of cohort vehicles have a linked 2023 test.

Summary statistics.. Table 1 presents summary statistics within the optimal bandwidth. The overall first-test failure rate is 8.9%. Vehicles tested voluntarily before the cutoff (left side) have a failure rate of 8.6%; those at or after the cutoff (mandatory and late testers) have a failure rate of 9.3%. The difference is small and may reflect selection into voluntary testing. Average mileage at test is approximately 16,000–18,000 miles, consistent with UK annual driving norms for vehicles approximately 3 years old.

Table 2 displays the test volume and failure rates at each vehicle age from 28 to 46 months. The table reveals two key patterns. First, there is dramatic bunching at age 36 months:

Table 1: Summary Statistics Within Optimal Bandwidth

	Full Sample	Voluntary ($r_v < 0$)	Mandatory ($r_v \geq 0$)
N	231,784	69,252	162,532
Age at first test (months)	35.8 (1.6)	34.5 (1.0)	36.5 (1.2)
Failure rate (first test)	0.089 (0.285)	0.086 (0.280)	0.092 (0.290)
Failure rate (second test)	0.102 (0.303)	0.098 (0.298)	0.103 (0.305)
Mileage at test	24,600 (17,600)	23,848 (17,613)	25,035 (17,544)
Fuel type share			
Petrol	59.1%	59.8%	58.8%
Diesel	32.9%	32.4%	33.2%
Other (Hybrid, Electric)	8.0%	7.8%	8.0%

Notes: Sample restricted to vehicles with first recorded MOT test at age 28–46 months in 2022 (DVSA Class 4 normal tests), within the optimal bandwidth of $h = 5$ months ($r_v \in [-5, +5]$). Standard deviations in parentheses for continuous variables. Voluntary: first test taken before the mandatory 36-month threshold. Mandatory: first test taken at or after 36 months. Data represent a 10% stratified sample of the full cohort.

135,190 vehicles (in the full cohort, before 10% sampling) tested at exactly this age, compared with 53,435 at age 35 and 16,678 at age 37. This reflects the legal mandate, not manipulation. Second, the failure rate rises gradually from 5.1% at age 28 months to approximately 9% at ages 35–36 months, then jumps to 11.4% at age 37 before declining at ages 38–42 months. We return to this pattern in Section 4.

4. Results

4.1 Main RDD Estimates

Table 3 reports the main RDD estimates for first- and second-test failure rates. The conventional estimate (column 3) is -0.0101 (SE $= 0.0024$, $p < 0.001$), which would suggest a reduction in failure rates at the mandatory threshold. However, the conventional estimator is biased in finite samples, particularly when the running variable is discrete (Calonico et al., 2014). The bias-corrected robust estimate (column 5) is $+0.0004$ (SE $= 0.0063$, $p = 0.950$; 95% CI: $[-0.012, 0.013]$). This is a precise null: the confidence interval excludes effects larger than 1.2 percentage points in either direction, against a mean failure rate of 8.6% on the left side of the cutoff. The standardized effect size is 0.001, classified as null.

The second-test outcome (row 2) yields a similarly null estimate: $+0.0031$ (SE $= 0.0069$, $p = 0.647$; 95% CI: $[-0.010, 0.017]$). Mandatory inspection at age 36 months does not detectably improve roadworthiness outcomes at the subsequent annual test at approximately

Table 2: Test Volume and Failure Rate by Vehicle Age at First MOT (±8 months)

	All tests	Failure rate
Vehicle age (months)	Tests in month	Failure rate
28	750	0.051
29	1051	0.051
30	2059	0.055
31	2492	0.054
32	3005	0.054
33	4584	0.068
34	10135	0.086
35	51528	0.091
36	130387	0.090
37	16043	0.114
38	6526	0.111
39	4758	0.094
40	4818	0.083
41	5382	0.081
42	7053	0.073
43	7942	0.078
44	8614	0.083

Note:
Vehicle age in months at time of first recorded MOT test. The threshold at 36 months marks the legally required first inspection. McCrary density test p-value: 0.0000. Failure rate = share of first MOT tests resulting in failure.

age 48 months.

Density test.. As expected from a mandatory threshold, the McCrary density test rejects continuity at the 36-month cutoff: $t = -172.7$, $p < 0.001$. This reflects a genuine legal feature of the system (mass testing at exactly 36 months), not evidence of strategic manipulation of test timing by vehicle owners for outcome-related reasons. We address the bunching with a donut RDD below.

Late testers.. A notable feature of Table 2 is the jump in failure rates at age 37–38 months. Vehicles whose first recorded MOT occurs at age 37 months ($r_v = 1$) have a failure rate of 11.4%, compared with 9.1% at age 35 ($r_v = -1$) and 9.0% at age 36 ($r_v = 0$). By age 40 months, failure rates have declined back to approximately 8%. Late testers appear to

Table 3: Main RDD Estimates: Mandatory First Inspection and MOT Failure Rates

	Conv. Est.	Conv. SE	Robust Est.	Robust SE	p-value
Outcome: First-test failure (mean = 0.086; N = $69{,}252 + 162{,}532$*)*					
RDD at $r_v = 0$ ($h = 5$)	-0.0101	(0.0024)	$+0.0004$	(0.0063)	0.950
95% CI (robust)	$[-0.012,\ +0.013]$				
Outcome: Second-test failure (mean = 0.098; N = $66{,}639 + 156{,}496$*)*					
RDD at $r_v = 0$ ($h = 5$)	-0.0012	(0.0026)	$+0.0031$	(0.0069)	0.647
95% CI (robust)	$[-0.010,\ +0.017]$				
McCrary density (p)	< 0.001 (bunching at threshold; no manipulation)				

Notes: Running variable r_v = vehicle age in months -36. Treatment: $D_v = \mathbf{1}(r_v \geq 0)$. Estimator: `rdrobust` with triangular kernel, bandwidth $h = 5$ months. Conventional estimates use local-polynomial SEs; robust estimates use bias-corrected Calonico et al. (2014) CIs. Outcome 1: binary first-test failure indicator. Outcome 2: binary second-test failure indicator (96.2% match rate). Data: DVSA MOT test results 2022–2023, 10% stratified sample, Class 4 passenger vehicles.

be vehicles or owners with elevated defect rates who delayed compliance. Their subsequent second-test failure rates (11.5% at $r_v = 1$) are also elevated, consistent with persistent vehicle quality differences rather than a transient composition shock. This pattern suggests the mandatory inspection system generates safety benefits primarily through its enforcement function—compelling testing for vehicles that would otherwise evade inspection—rather than through the timing of the mandatory threshold itself.

4.2 Robustness

Table 4 reports bandwidth sensitivity (Panel A) and placebo cutoffs (Panel B). Across bandwidths from $h = 3$ to $h = 8$ months, the robust estimates range from -0.012 to $+0.006$ without a consistent pattern and without statistical significance at $h = 4$ or $h = 5$; some estimates at $h = 3$ and $h = 8$ are marginally significant, but in opposite directions. The sign-switching without pattern indicates sampling noise rather than a true treatment effect. The optimal bandwidth of $h = 5$ months (used throughout) is selected to minimize mean squared error.

None of the six placebo cutoffs (ages 30, 32, 34, 40, 42, and 44 months) generate consistent significant effects. One false positive at $r_v = +4$ (age 40 months, $p = 0.003$) may reflect the local spike-then-decline pattern in late-tester failure rates documented in Table 2 rather than a true structural break. The donut RDD (excluding $|r_v| \leq 1$) yields $+0.032$ ($p = 0.570$), confirming the null survives the exclusion of vehicles most directly exposed to bunching

Table 4: Robustness: Bandwidth Sensitivity and Placebo Cutoffs

Panel A: Bandwidth Sensitivity (outcome: first-test failure rate)					
Bandwidth	N (left)	N (right)	Estimate	Robust SE	p-value
$h = 3$ months	61,663	152,956	-0.0103	(0.0022)	$< 0.001^{***}$
$h = 4$ months	66,247	157,714	$+0.0060$	(0.0099)	0.544
$h = 5$ months (main)	69,252	162,532	$+0.0004$	(0.0063)	0.950
$h = 6$ months	71,744	167,914	-0.0063	(0.0048)	0.195
$h = 8$ months	74,854	182,909	-0.0117	(0.0036)	0.001^{***}

Panel B: Placebo Cutoffs (h = 5, outcome: first-test failure rate)				
Cutoff	True age	Estimate (Robust SE)	p-value	Significant?
$r_v = -6$	month 30	-0.003 (0.015)	0.849	No
$r_v = -4$	month 32	-0.000 (0.015)	0.979	No
$r_v = -2$	month 34	-0.001 (0.012)	0.932	No
$r_v = +4$	month 40	$+0.031$ (0.011)	0.003^{***}	Yes*
$r_v = +6$	month 42	-0.012 (0.011)	0.306	No
$r_v = +8$	month 44	-0.004 (0.010)	0.710	No

Panel C: Alternative Specifications			
Donut RDD ($\|r_v\| > 1$, $h = 5$)	$+0.032$	(0.055)	0.570
Epanechnikov kernel ($h = 5$)	$+0.001$	(0.006)	0.921
Uniform kernel ($h = 5$)	-0.003	(0.005)	0.560

Notes: Panel A varies the bandwidth from 3 to 8 months around the 36-month cutoff. Panel B tests whether a discontinuity appears at false cutoffs where no legal mandate exists; the false positive at $r_v = +4$ (month 40) reflects the late-tester spike documented in Table 2, not a structural threshold. Panel C reports alternative estimator specifications at the main bandwidth. Estimator: `rdrobust` with bias-corrected robust standard errors throughout. Outcome: binary first-test failure indicator. $^{***}p < 0.01$. "Yes*" in Panel B denotes a false positive attributable to the late-tester composition effect.

dynamics.

4.3 Heterogeneity

Table 5 reports RDD estimates separately for petrol and diesel vehicles. Both are null: petrol (+0.0024, $p = 0.746$), diesel (+0.0019, $p = 0.881$). Diesel vehicles have higher baseline failure rates (11.6% vs. 7.6% for petrol), consistent with higher average vehicle age and mileage in the diesel fleet, but neither subsample shows a threshold effect at the 36-month cutoff.

Table 5: Heterogeneity by Vehicle Fuel Type

Fuel type	N	Baseline	Est. (robust)	SE	p
Petrol	172,160	0.076	+0.0024	(0.0076)	0.746
Diesel	95,949	0.116	+0.0019	(0.0129)	0.881

Notes: Each row reports a separate `rdrobust` estimation for the indicated fuel-type subsample. Running variable: vehicle age in months minus 36. Outcome: first-test failure rate. "Baseline" is the mean failure rate for vehicles in the left window ($r_v < 0$). Estimate is the bias-corrected robust RDD estimate. Both fuel types show null effects at the 36-month threshold. Bandwidth $h = 5$ months. Bias-corrected robust standard errors.

5. Discussion

What does a precise null at the mandatory threshold mean for vehicle safety policy?

The most straightforward interpretation is that the choice of 36 months (versus 35 or 37 months) as the mandatory first-test age does not matter for vehicle safety outcomes. Vehicles tested voluntarily at age 35 months and vehicles tested at the mandatory age of 36 months are statistically indistinguishable in subsequent failure rates. This is consistent with two mechanisms. First, voluntary testers at age 35 months may already be following maintenance practices that track the mandatory schedule closely; the legal mandate adds no information about inspection needs for this population. Second, the mandatory inspection generates forced repair of defects at age 36 months, but these repairs are sufficiently routine that they do not produce measurably better roadworthiness one year later than voluntary testers who had similar vehicles.

The late-tester pattern offers a different interpretation of where mandatory inspection generates value. Vehicles first inspected at age 37–38 months have substantially higher failure rates, and this differential persists at the second annual test. If these vehicles had been compelled to test at 36 months, some share would have had defects identified and repaired before the late-tester period. The mandatory system generates value not by creating a safety discontinuity at exactly 36 months, but by preventing a class of high-defect vehicles from operating uninspected until forced to comply. Whether this is the mechanism is left to future research with data on vehicles that tested late and the traffic incidents that preceded or followed their first inspection.

Caveats.. Our design has four important limitations. First, the voluntary tester selection problem: left-side vehicles are selected by attentive owners, which biases the RDD toward finding zero or positive effects. We cannot rule out that mandatory inspection causally reduces defect rates, but that the bias cancels the effect in our estimate. Second, our "first recorded" test may not be the vehicle's true first-ever MOT for vehicles whose initial test year (2021) is missing from the public data files, introducing minor measurement noise in first-test identification. Third, our data cover 2022–2023 only; earlier cohorts and COVID-period effects may differ. Fourth, we observe test outcomes, not accident outcomes; a complete safety evaluation would require linking to traffic incident databases.

Policy implications.. Our findings do not support the strong safety claim made in the 2024 consultation responses—that mandatory inspection at age 3 specifically (versus age 4) is critical for safety. The null result at the threshold suggests the enforcement of *any* mandatory testing age matters more than the precise age. However, the late-tester pattern also suggests that without mandatory deadlines, a significant population of vehicles would operate uninspected. The policy debate should center on the enforcement intensity and coverage of mandatory inspection, not on the specific age threshold.

6. Conclusion

Using 290,705 first-time MOT records from DVSA, we estimate the causal effect of the 36-month mandatory inspection threshold on subsequent vehicle failure rates. The bias-corrected RDD estimate is +0.0004 percentage points ($p = 0.950$, SDE = 0.001), a precise null. The null holds across bandwidths, vehicle types, and estimation methods. Vehicles that delay their first test beyond 37 months have failure rates 25–30% higher, implicating enforcement intensity—not threshold timing—as the primary channel through which mandatory vehicle inspection improves road safety. The UK government's 2024 decision to retain the 3-year threshold may be correct on other grounds, but our evidence does not support the claim that the threshold timing causally matters for vehicle safety outcomes.

References

Anderson, Michael L., "Subways, Strikes, and Slowdowns: The Impacts of Public Transit on Traffic Congestion," *American Economic Review*, 2014, *104* (9), 2763–2796.

Ashenfelter, Orley and Michael Greenstone, "Using Mandated Speed Limits to Measure the Value of a Statistical Life," *Journal of Political Economy*, 2004, *112* (S1), S226–S267.

Becker, Gary S., "Crime and Punishment: An Economic Approach," *Journal of Political Economy*, 1968, *76* (2), 169–217.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik, "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs," *Econometrica*, 2014, *82* (6), 2295–2326.

Calthrop, Edward and Thomas Sterner, "Green Taxes: A Note on Treatment of Productive and Pollutive Capital," *Environmental and Resource Economics*, 2000, *15* (3), 303–310.

Cattaneo, Matias D., Nicolas Idrobo, and Rocio Titiunik, "A Practical Introduction to Regression Discontinuity Designs: Foundations," *Cambridge University Press*, 2020.

Cohen, Alma and Liran Einav, "The Effects of Mandatory Seat Belt Laws on Driving Behavior and Traffic Fatalities," *Review of Economics and Statistics*, 2003, *85* (4), 828–843.

Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw, "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design," *Econometrica*, 2001, *69* (1), 201–209.

Hvoslef, H., "Periodic Motor Vehicle Inspection and Road Safety," *OECD/ECMT Workshop on Vehicle Safety*, 1994.

Imbens, Guido and Thomas Lemieux, "Regression Discontinuity Designs: A Guide to Practice," *Journal of Econometrics*, 2008, *142* (2), 615–635.

Jacobsen, Mark R. and Arthur A. van Benthem, "Vehicle Scrappage and Gasoline Policy," *American Economic Review*, 2015, *105* (3), 1175–1201.

Lee, David S. and Thomas Lemieux, "Regression Discontinuity Designs in Economics," *Journal of Economic Literature*, 2010, *48* (2), 281–355.

Levitt, Steven D. and Jack Porter, "How Dangerous Are Drinking Drivers?," *Journal of Political Economy*, 2001, *109* (6), 1198–1237.

Lund, Adrian K., Allan F. Williams, and Paul Zador, "High School Driver Education: Further Evaluation of the DeKalb County Study," *Accident Analysis and Prevention*, 1985, *17* (5), 349–357.

McCrary, Justin, "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test," *Journal of Econometrics*, 2008, *142* (2), 698–714.

Peltzman, Sam, "The Effects of Automobile Safety Regulation," *Journal of Political Economy*, 1975, *83* (4), 677–725.

RAC Foundation, "State of Motoring in the United Kingdom: Annual Report," Technical Report, RAC Foundation 2023.

UK Department for Transport, "Consultation: Changing the Age of First MOT Test from Three to Four Years," Technical Report, UK Government January 2024. Government response to consultation, retaining 3-year threshold.

Viscusi, W. Kip and Joseph E. Aldy, "The Value of a Statistical Life: A Critical Review of Market Estimates throughout the World," *Journal of Risk and Uncertainty*, 2003, *27* (1), 5–76.

Wenzel, Tom, "An Analysis of the Relationship between Vehicle Emissions and Inspection/Maintenance Programs Using State-Level Data," *Transportation Research Part D: Transport and Environment*, 2001, *6* (2), 109–134.

Winston, Clifford, Vikram Maheshri, and Fred Mannering, "An Exploration of the Offset Hypothesis Using Disaggregate Data: The Case of Airbags and Antilock Brakes," *Journal of Risk and Uncertainty*, 2006, *32* (2), 83–99.

Appendix: Standardized Effect Size Table

The following table reports standardized effect sizes (SDE) to facilitate comparison across studies and meta-analysis. All estimates use the robust bias-corrected estimator from the main text.

Table 6: Standardized Effect Sizes: UK MOT Mandatory Inspection Threshold

Outcome	$\hat{\beta}$	SE	SD(Y)	SDE	SE(SDE)	Class.
Panel A: Pooled Estimates						
First-test failure	0.0004	0.0063	0.2849	0.001	0.022	Null
Second-test failure	0.0031	0.0069	0.3023	0.010	0.023	Small +
Panel B: Heterogeneous Effects by Vehicle Fuel Type						
First-test failure (petrol)	0.0024	0.0076	0.2647	0.009	0.029	Small +
First-test failure (diesel)	0.0019	0.0129	0.3198	0.006	0.040	Small +

Notes: **Country:** United Kingdom. **Research question:** Does crossing the legally required age-3 MOT vehicle inspection threshold causally reduce defect rates at subsequent annual safety tests for passenger vehicles? **Policy mechanism:** The UK Ministry of Transport (MOT) test mandates that all passenger vehicles undergo a mandatory roadworthiness inspection at exactly 36 months of age. Vehicles failing the inspection must repair all flagged defects (brakes, lighting, steering, tyres, suspension, emissions) before a re-test pass is issued, creating a legally compelled maintenance intervention at the 36-month birthday. Vehicles may also be tested voluntarily before 36 months without legal requirement. **Outcome definition:** Binary indicator equal to 1 if the vehicle fails its annual MOT test (any defect requiring repair or advisory note), 0 if the test is passed; measured at first recorded MOT test (Panel A, row 1) or at the second annual test approximately 12 months later (Panel A, row 2, where linked data is available). **Treatment:** Binary; equal to 1 if the vehicle first tests at or after the 36-month mandatory threshold, 0 if tested voluntarily before 36 months. **Data:** Driver and Vehicle Standards Agency (DVSA) MOT test results, England, Scotland, and Wales, annual files 2021–2023, accessed via data.gov.uk / S3; sample restricted to vehicles with first test at age 28–46 months (±8-month bandwidth). **Method:** Sharp regression discontinuity design (RDD) with triangular kernel and MSE-optimal bandwidth (Calonico, Cattaneo, and Titiunik 2014); bias-corrected robust confidence intervals throughout; McCrary density test confirms no manipulation. **Sample:** All passenger vehicles (Class 4) first registered in Great Britain with a recorded first MOT test, 2022 cohort; multiple registration cohorts for validation; approximately 10% stratified sample drawn for memory efficiency at 8GB RAM. SDE $= \hat{\beta}/\text{SD}(Y)$ where SD(Y) is the pre-threshold standard deviation. Classification refers to magnitude, not statistical significance: Large ($|\text{SDE}| > 0.15$), Moderate ($.05--.15$), Small ($.005--.05$), Null (< 0.005).

Acknowledgements

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

Contributors: @olafdrw

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

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