Online Appendix

Re-examining Moral Hazard in Risky Driving:

New Evidence from Behavioral Data in Auto Insurance

Yizhou Jin

July 12, 2026

D Robustness Analyses

This appendix collects the robustness analyses supporting each of the paper's designs—the hazard model, the precipitation response, the deductible test, and the text-message experiment—followed by the data-and-algorithm appendix.

D.1 HPU Risk Estimation Robustness

This subsection supports Section 3: robustness of the hazard-model estimates of HPU's accident contribution.

D.2 HPU Response to Precipitation Robustness

This subsection supports Section 4.1 (precipitation shocks): a fleet-reweighted replication of the main precipitation panel, a Poisson QMLE check, the within-trip interval-level panel from Dataset 1, and the full set of control coefficients across the paper's canonical specifications. The back-of-envelope calibration of the implied moral-hazard elasticity $\hat{\beta}$

Table D1: Sensitivity Analysis: Effect of Interval Width

Interval Width	N	HPU OR	Precip. OR	HPU × Precip. OR
5 sec	12,965,344	2.029***	1.998***	1.497**
10 sec	6,494,874	1.972***	1.961***	1.643***
30 sec	2,181,248	1.811***	1.906***	1.609**

Notes: $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$ (on the log-odds scale); standard errors delta-method. HPU OR is the main effect (dry, daytime, mean speed); Precip. OR and HPU × Precip. OR are per inch/hr. Five-second intervals are aggregated to coarser widths by backward bucketing from the accident interval; each width re-estimates the Table 2 column-(7) specification.

Table D2: Robustness: Sampling Design and Weighting

	(A) Main		(B) No Controls		(C) 2 Controls		(D) IPSW	
	Est.	(SE)	Est.	(SE)	Est.	(SE)	Est.	(SE)
HPU	0.7077***	(0.0927)	1.2716***	(0.0855)	0.7697***	(0.0987)	0.7045***	(0.0980)
Speed (mph)	0.0086***	(0.0008)	0.0155***	(0.0009)	0.0100***	(0.0009)	0.0089***	(0.0009)
Precipitation	0.6921***	(0.0785)	0.6748***	(0.0824)	0.6701***	(0.0854)	0.7085***	(0.0785)
Night	0.4332***	(0.0377)			0.6165***	(0.0457)	0.4891***	(0.0427)
HPU × Precip	0.4033**	(0.1716)	0.3262	(0.2184)	0.4849**	(0.1946)	0.4619**	(0.1893)
HPU × Speed	0.0026	(0.0018)	0.0023	(0.0022)	0.0037*	(0.0020)	0.0030	(0.0020)
Speed × Precip	0.0021	(0.0021)	−0.0018	(0.0025)	0.0005	(0.0022)	0.0031	(0.0022)
HPU × Night	0.3270***	(0.1193)			0.2998**	(0.1302)	0.3601***	(0.1259)
Observations	12,965,344		840,152		3,269,882		12,965,344	
Trips	48,661		4,428		13,284		48,661	
HPU OR (mean cond.)	2.37		3.64		2.50		2.40	
FE	Driver		Trip		Driver		Driver	

Notes: $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$. Logit coefficients (log odds ratios); standard errors in parentheses. All columns use the col. (7) specification from Table 2; precipitation enters in inches per hour. Column (A) is the main specification: 10 control trips per driver, unweighted, driver FE. Column (B) uses accident trips only with trip FE (night, weekend, season absorbed); two accident trips enter with a single five-second interval and are dropped by the trip FE. Column (C) randomly subsamples to 2 control trips per driver (drivers with fewer contribute all of theirs). Column (D) applies IPSW weights that upweight each control trip to represent all non-accident trips by the same driver. Slope coefficients are consistent under case-control sampling without reweighting (Prentice and Pyke 1979); the similarity of columns (A) and (D) confirms this.

from each precipitation specification and from the insurance experiment is documented in the paper's Appendix C.

Control coefficients across canonical specifications. Table D6 consolidates the control-variable coefficients from the full specification of each of the paper's three trip-level designs (the precipitation LPM and offset-Poisson in the avg-intensity form, Table 3 columns 4 and 8, and the experimental DiD Poisson, Table 5 column 3, the canonical experiment specification), whose union of control vectors is the fifteen harmonized trip-level controls.

Table D3: Trip-level Dataset Sum Stats: Fleet-Representative Reweighting (Robustness)

Variable	Mean	Std. Dev.	Pctl(10)	Median	Pctl(90)
Driver-level					
Age (years)	43.1	12.6	28.1	41.4	61.4
Is female	0.13	0.34	0.00	0.00	1.00
Tenure (years)	1.7	1.2	0.4	1.4	3.4
Past rides completed	734.8	513.9	84.0	704.0	1,457.0
Trip-level					
Distance (miles)	6.2	5.8	1.7	4.5	12.7
Duration (minutes)	17.0	9.1	8.6	15.1	27.8
Pct. drive-time with passenger	0.72	0.14	0.52	0.74	0.90
Speed (mph)	19.3	9.2	9.2	17.4	32.5
Is weekend	0.30	0.46	0.00	0.00	1.00
Is night-time	0.31	0.46	0.00	0.00	1.00
Is raining	0.27	0.44	0.00	0.00	1.00
Rain intensity (in/hr | rainy)	0.074	0.382	0.000	0.000	0.075
Is snowing	0.08	0.28	0.00	0.00	0.00
Handheld phone use (m/km)	8.6	32.2	0.0	0.0	21.4
Harsh braking per 1k miles	186.4	389.9	0.0	0.0	562.9
Harsh cornering per 1k miles	29.78	121.82	0.00	0.00	56.53

Notes: Robustness version of Table 1 with fleet-representativeness weights $w_i = 1/p_{s(i)}$, the inverse of the stratum sampling probabilities (Section 2.2). Because the design oversamples high-HPU drivers, the weights restore the low-HPU drivers thinned from the sample and recover a broader Lyft fleet snapshot (approximately 16 million driver-trips). The weighted moments are fleet-representative on driver age, gender, past rides completed, trip distance, duration, passenger share and speed, the weekend, night, and rain shares, harsh braking, and handheld phone use. All reduced-form analyses in the main text use the unweighted sample.

D.3 Robustness of the Deductible Test

This subsection supports Section 4.2. Section D.3.1 sets out the dose's properties and why the twelve contaminated states are excluded; Section D.3.2 then stress-tests both specifications in a consolidated exhibit and two measurement checks. All deliver the same two-part message as the main analysis: within-state variation gives a precise null and across-state variation a small negative response significant only on the extensive margin, both far below the magnitude implied by classical moral hazard.

D.3.1 Properties of the dose

The deductible test's dose, ΔEOOP_c, is derived in Section 4.2 from the insurer's county expected collision costs at the \$500 and \$1,000 deductibles under the 2021 pricing, which is built on pre-2021 data; the dose is therefore a predetermined, county-level characteristic, fixed within driver, and the design does not rely on within-driver dose variation. Figure D1

Table D4: Precipitation Effects on HPU: Fleet-Reweighted (Robustness)

	Extensive: P(HPU >0) (LPM)				QMLE: log E[HPU] (Poisson)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Rain indicator	−0.0064	−0.0050	−0.0041	−0.0042	−0.0244	−0.0222	−0.0151	−0.0145
	(0.0081)	(0.0075)	(0.0044)	(0.0043)	(0.0309)	(0.0323)	(0.0328)	(0.0328)
Rain × start intensity	0.0371	0.0331	−0.0201*		0.0072	−0.0162	−0.0633	
	(0.0226)	(0.0206)	(0.0106)		(0.0563)	(0.0576)	(0.0499)	
Rain × Δ intensity	−0.0079	−0.0071	−0.0239		−0.1235	−0.1462	−0.1444	
	(0.0439)	(0.0435)	(0.0291)		(0.1191)	(0.1195)	(0.0934)	
Rain × avg intensity				−0.0193*				−0.0619
				(0.0106)				(0.0499)
%Δ HPU at mean rain	2.181	1.944	−1.179*	−1.099*	0.133	−0.299	−1.166	−1.103
	(1.328)	(1.208)	(0.623)	(0.601)	(1.045)	(1.064)	(0.914)	(0.884)
Location + Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Trip-level controls (15)	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Driver FE	No	No	Yes	Yes	No	No	Yes	Yes
Observations	1,689,022	1,689,022	1,687,611	1,687,611	1,677,339	1,677,339	1,675,350	1,675,350

Notes: $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$. This table re-estimates the *identical* specifications, controls, and fixed effects as Table 3 (extensive LPM in cols 1–4; exposure-offset Poisson QMLE in cols 5–8), with every fit weighted by the inverse stratum sampling probability so the estimation sample is HPU-representative of the platform fleet. Standard errors (in parentheses) are two-way clustered by driver and date. Fleet-weighted baseline $\Pr(\text{HPU} > 0) = 0.315$; the %Δ HPU row is evaluated at the fleet-weighted baseline and rainy-trip mean intensity, consistent with the table's weighting. Because the sample oversamples high-HPU drivers, the inverse sampling-probability weights upweight the relatively few low-HPU drivers and concentrate weight on a small effective sample ~~(a Kish effective N of roughly 900)~~. The within-driver extensive estimate (cols 3–4) is moral-hazard-signed and several times the unweighted magnitude, marginally significant; the no-driver-FE columns (cols 1–2) flip positive and the Poisson columns (cols 7–8) are insignificant, reflecting the small effective sample. We therefore read the fleet *level* cautiously. Its *direction* is informative: the response is larger once the low-HPU drivers carry their fleet weight, because those drivers are the most responsive to everyday incentives (Section 5). The fleet average also targets a different estimand from the main text, which reports the response among the HPU-active drivers where the behavior and its externality concentrate. %Δ HPU at mean rain follows the same delta-method formula as Table 3.

maps the identifying variation across the estimation sample's counties.

The twelve excluded states (the Figure D1 note gives the selection criteria) are dropped because their shocks sit inside the event window—the storm states' recovery occupies the pre-period, the policy states' rollouts the post-period—so no window trim can rescue the affected states. Re-adding them under the storm-extended pricing fails the design's pre-trend flags (Table D7, "+ storm states" row).

D.3.2 Robustness by dimension

The two identified specifications of Section 4.2—within-state (State×Date) and across-state (Date)—answer differently: within-state variation in expected out-of-pocket exposure produces a precise null, while across-state variation produces a small negative reduction that is statistically significant only on the extensive margin. We stress-test this two-part pattern in a single consolidated exhibit (Table D7).

Table D5: Precipitation Effects on HPU

	Logit			LPM		
	(1)	(2)	(3)	(4)	(5)	(6)
Precipitation (in/hr)	0.0061	−0.0105	−0.0359	0.000231	−0.000400	−0.001198***
	(0.0077)	(0.0078)	(0.0365)	(0.000292)	(0.000293)	(0.000277)
%Δ HPU at mean rain	0.12	−0.21	−0.73	0.12	−0.21	−0.62
Implied $\hat{\beta}$	0.004	−0.006	−0.021	0.003	−0.006	−0.018
Controls (speed, night)	No	Yes	Yes	No	Yes	Yes
Driver FE	No	No	Yes	No	No	Yes

Notes: $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$. Standard errors in parentheses. Dependent variable is the five-second HPU indicator; all six columns use the full sample of 12,965,344 accident-driver intervals. Columns (1)–(3) are logistic regressions; columns (4)–(6) are linear probability models, with the driver fixed effects in columns (3)/(6) fit by conditional logit / within-driver demeaning. Controls are speed (mph) and a night indicator. Precipitation enters in inches per hour, so its coefficients are per inch/hr. %Δ HPU at mean rain evaluates each coefficient at the rainy-trip mean intensity ($x^* = 0.203$ in/hr): LPM columns use $100\,\hat{\beta}\,x^*$/base (base $P(\text{HPU}) = 0.039$), logit columns $100(\exp(\hat{\beta}\,x^*) - 1)$. Implied $\hat{\beta}$ is the moral-hazard elasticity (%ΔHPU)/(%ΔMC), with %ΔMC = 34.7% the marginal-cost increase from Table 2 column 7 (driver-FE hazard) evaluated at the same x^*.

Consolidated robustness. Table D7 reports both specifications side by side across the inference level (county-, driver-, and state-clustered standard errors), the sample (adding the twelve excluded states back under the storm-extended pricing; narrowing the window to ±8 days), the dose construction (a Pareto severity family in place of the exponential; the 2023 pricing vintage in place of 2021), and an alternative labor-supply outcome (driving hours). Each cell carries a two-part parallel-trends flag (extensive, intensity) evaluated at the 5% level. The within-state estimate is indistinguishable from zero in every cell and on both margins; the across-state extensive reduction ($\approx -0.91\%$) recurs at a similar magnitude under both severity families and both pricing vintages—while the across-state intensity margin stays indistinguishable from zero in every county- and state-clustered storm-excluded cell (the driver-clustered standard error, much smaller on this margin, is the exception, which is why our preferred inference clusters at the county grain of the dose); the storm re-inclusion row fails the pre-trend flags, consistent with Section D.3.1.

D.4 Experimentation Robustness

This subsection supports Section 5: an illustration of the text-message intervention used in the field experiment, and the full-set robustness of the average treatment effects across outcomes (HPU, harsh acceleration, hours driven) and progressive control specifications.

Table D6: Trip-level Control Estimates

	(1) Precipitation T.3 col. 4 LPM $\Pr(\text{HPU} > 0)$	(2) Precipitation T.3 col. 8 Poisson $\log E[\text{HPU}]$	(3) Experiment T.5 col. 3 Poisson $\log E[\text{HPU}]$
Harmonized trip-level controls (*trip economics, origin demand, in-app interactions*)			
Driver earnings ($/mi)	0.0034^{***}	0.0026^{**}	-0.0632^{***}
	(1.97×10^{-4})	(0.0012)	(0.0032)
Pickup-phase (P2) duration share	0.1347^{***}	1.1345^{***}	—
	(0.0036)	(0.0218)	
Origin population density (per sq. mi)	$1.55 \times 10^{-7*}$	$-7.81 \times 10^{-7**}$	$-3.78 \times 10^{-6***}$
	(8.01×10^{-8})	(3.68×10^{-7})	(6.61×10^{-7})
Origin ride requests (per hour)	$-9.23 \times 10^{-5***}$	$-5.84 \times 10^{-4***}$	-0.0019^{***}
	(1.68×10^{-5})	(7.66×10^{-5})	(2.27×10^{-4})
Rides added (per mi)	-0.0625^{***}	0.1952^{***}	0.5628^{***}
	(0.0023)	(0.0144)	(0.0340)
Rides canceled (per mi)	0.0698^{***}	-0.1501^{***}	abs.
	(0.0055)	(0.0269)	
Rider cancellations (per mi)	4.02×10^{-4}	0.0269^{*}	-0.0360
	(0.0034)	(0.0148)	(0.0578)
Pickups (per mi)	-3.10×10^{-4}	-0.0797^{***}	0.1414^{***}
	(0.0024)	(0.0104)	(0.0163)
Dropoffs (per mi)	-1.66×10^{-4}	0.1414^{***}	-0.0716^{*}
	(0.0049)	(0.0225)	(0.0385)
Reroutes (per mi)	4.49×10^{-4}	3.95×10^{-4}	0.0233^{***}
	(3.99×10^{-4})	(0.0016)	(0.0085)
Calls from rider (per mi)	0.0461^{***}	0.0847^{***}	0.2065^{***}
	(0.0017)	(0.0079)	(0.0132)
Texts from rider (per mi)	0.0304^{***}	0.0567^{***}	0.2543^{***}
	(0.0024)	(0.0104)	(0.0323)
Platform pushes (per mi)	0.0055^{***}	0.0098^{***}	0.1400^{***}
	(7.03×10^{-4})	(0.0025)	(0.0422)
Texts from platform (per mi)	0.0443^{***}	0.0959^{***}	0.3378^{***}
	(0.0039)	(0.0156)	(0.0528)
App crashes (per mi)	0.0248	-0.2206	-2.49×10^{3}
	(0.0316)	(0.1550)	(2.08×10^{3})
Exposure offset	—	log(miles)	log(miles)
Location-pair FE	Yes	Yes	—
Date FE	Yes	Yes	Yes
Hour FE	Yes	Yes	Yes
Driver FE	Yes	Yes	Yes
SE cluster	Driver × date	Driver × date	Driver
Observations	1,687,611	1,675,350	463,872

Notes: $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$; standard errors in parentheses, clustered as stated in the SE-cluster row. Each column reports the control-variable coefficients from the full specification of one of the paper's three trip-level designs (three different analyses on different samples); the row block is the union of their control vectors. Cols. (1)/(2) = Table 3 columns 4/8: trip-level extensive-margin LPM on $\Pr(\text{HPU} > 0)$ / Poisson QMLE on HPU with a log-mileage offset, avg-intensity, all 15 trip-level controls. Col. (3) = Table 5 column 3, the canonical experimental specification: trip-level DiD Poisson QMLE on HPU-miles with a log-mileage offset and driver, date, and hour fixed effects; per-mile controls winsorized at the 99th percentile; the pickup-phase share is not in the experiment's control set (its passenger margin is analyzed in columns 5–8 there). '—' = not part of that column's specification; 'abs.' = in the specification but dropped as collinear: in column (3) the 'Rides canceled' count is identically zero after winsorization (under 1% of segments nonzero). Trip distance enters only through the exposure offset, never as a linear control. Observations are the fits' estimation samples, slightly below the source-table panel size (464,045 trips): the Poisson likelihood drops missing-outcome/zero-exposure rows and all-zero-outcome fixed-effect cells. Auto-generated by `replication/code/30_precip/35_control_coefs.R`; do not edit by hand.

Robustness to the HPU outcome definition. The experimental treatment effect is robust to how HPU is measured. Table D9 re-estimates the combined main specification (Table 5) on the *extensive margin*—whether any handheld phone use occurs in the segment—and Table D10 with a *linear* model for HPU intensity in place of the Poisson estimator. Both reproduce the significant average treatment effect.

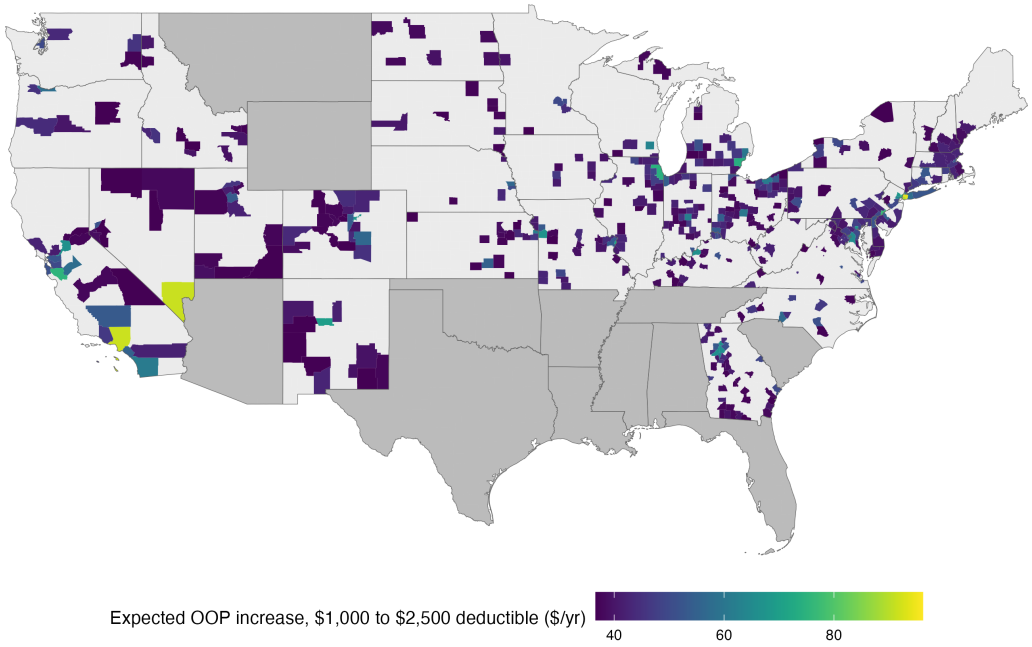

Figure D1: The deductible dose: county expected out-of-pocket increase

Notes: County expected annual out-of-pocket increase ΔEOOP_c ($/yr) for the $1,000→$2,500 deductible hike, derived from the insurer's county expected collision costs at the $500 and $1,000 deductibles under the exponential severity family (Section 4.2). Grey shaded states are the twelve excluded; unfilled counties are outside the estimation sample. The exclusions are selected ex ante on contemporaneous shocks inside the event window: the Winter-Storm-Uri FEMA major-disaster states (TX, OK, LA) and their affected neighbors (AR, MS, AL, TN), plus states with COVID-era policy changes in the same weeks—statewide mask-mandate drops (WY, MT) and capacity-cap relaxations (AZ, FL, SC). States with neither a FEMA declaration nor a February–March policy change remain in the sample.

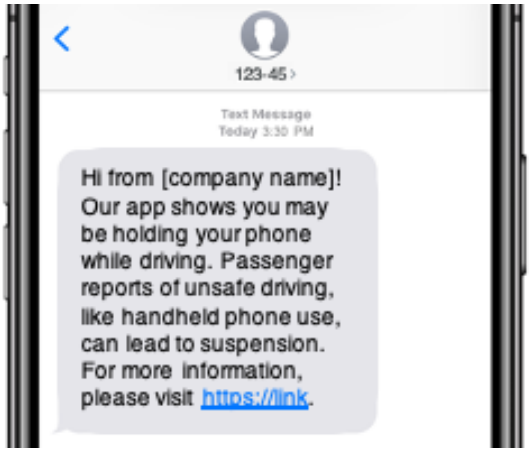

Figure D2: Illustration of the Text-Message Intervention

Notes: The text-message intervention used in our direct experiment (Section 5). Drivers in the treatment group received a text message highlighting recent HPU behavior and warning of potential consequences. The app interface shown may not represent the actual application.

Table D7: Deductible dose–response: consolidated robustness

	Within-state (Driver + State×Date)			Across-state (Driver + Date)		
Variant	Ext. %Δ	Int. %Δ	PT	Ext. %Δ	Int. %Δ	PT
Baseline	+0.20 (0.78)	+1.98 (1.91)	✓✓	−0.91** (0.42)	−3.42 (3.23)	✓✓
Clustering						
driver-clustered SE	+0.20 (0.77)	+1.98 (1.92)	✓✓	−0.91* (0.47)	−3.42*** (1.07)	✓✓
state-clustered SE	+0.20 (0.61)	+1.98 (2.39)	××	−0.91** (0.43)	−3.42 (2.98)	✓✓
Sample						
+ storm states	−0.16 (0.39)	+1.47 (1.70)	✓×	−0.83*** (0.29)	−5.34*** (1.65)	××
±8-day window	+0.04 (0.82)	+1.90 (1.92)	✓✓	−0.79* (0.44)	−2.92 (2.72)	✓✓
Dose construction						
Pareto severity	+0.10 (0.67)	+1.49 (1.66)	✓✓	−0.77** (0.31)	−2.96 (2.72)	✓✓
2023 pricing	+0.19 (0.65)	+1.48 (1.56)	✓✓	−0.80** (0.36)	−2.37 (2.81)	✓✓
Labor supply						
Driving hours	−1.31 (1.64)	−0.04 (2.24)	××	−1.50 (1.16)	+2.01 (2.03)	××

Notes. Each cell is the deductible dose–response as a percent change relative to the outcome's pre-period mean. *Ext.*: linear-probability model for $\mathbf{1}\{\text{outcome} > 0\}$ on Post×ΔEOOP, controlling for $\log(\text{driving miles})$ and passenger-phase share; $\%\Delta = 100\,\hat{\theta}\,\bar{x}/\bar{p}$. *Int.*: Poisson QMLE with a $\log(\text{driving-miles})$ offset; $\%\Delta = 100(e^{\hat{\theta}\,\bar{x}} - 1)$. Standard errors (parentheses) are clustered by county unless a row says otherwise; $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. *PT* flags the parallel-trends check (extensive, intensity): ✓ if the pre-period dose × event-time interactions are jointly not rejected at 5%, × if rejected (county-clustered Wald). The within-state columns add State×Date fixed effects (identifying θ from within-state dose differences); the across-state columns add Date fixed effects (also drawing on across-state variation); both include driver fixed effects and the controls above. Sample: driver-day panel, ±10-day window around March 1, 2021 with the twelve excluded states removed (Section 4.2) and the county expected out-of-pocket increase ΔEOOP (exponential severity, 2021 pricing; Section 4.2) in \$100/yr units as the dose, except where a row states otherwise. The severity-family and vintage rows rebuild the dose from the same pricing points. Labor supply: the extensive margin is $\mathbf{1}\{\text{drove}\}$ on a balanced within-span driver-day panel, the intensity is driving hours conditional on driving. Source: telematics driver-day panel; county pricing points, insurer pricing system; FARS fatal-crash counts, NHTSA.

Robustness to the control group. Our main specification identifies the treatment effect from a difference-in-differences that uses the full set of untreated targeted drivers as the comparison. Table D11 re-estimates the canonical specification (Table 5, column 3) against only the within-window randomized control group—drivers who logged on during the intervention window and were randomized out of treatment, the clean experimental contrast. The difference-in-differences and the randomized-control estimate agree on the intensity margin and are of the same sign and significance on the extensive margin, where the strict estimate is modestly larger, and the pre-period leads are jointly flat in both, so we adopt the higher-powered full-control difference-in-differences as our main specification.

Robustness of the hours-driven margin. The hours-driven coefficient is a precise null in the canonical specification (Table D8, hours panel, column (3): $+0.0010$ hours per trip, insignificant), and no column of the hours panel is statistically significant. A driver + day + hour fixed-effects difference-in-differences without the harmonized controls returns a

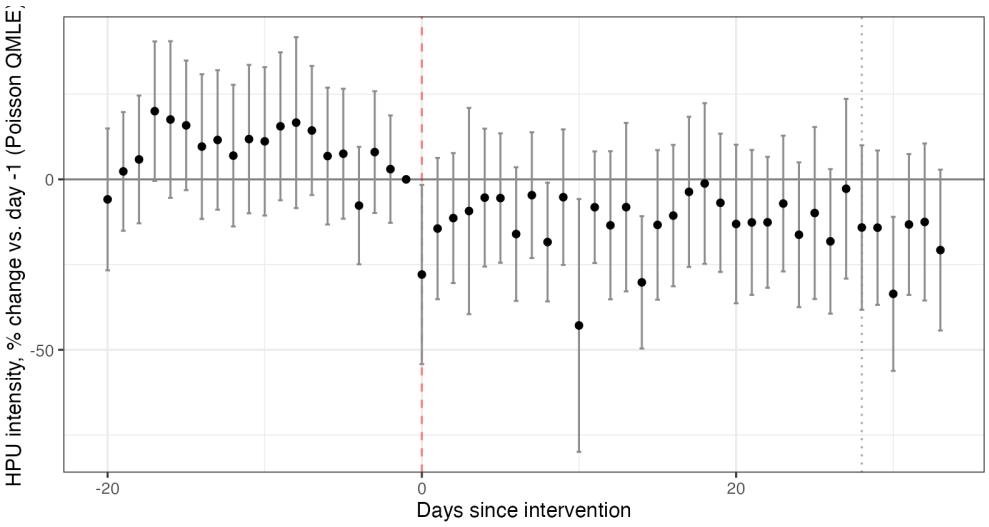

Figure D3: Direct Intervention: Day-by-Day Treatment Effect on HPU

Notes: Day-by-day disaggregation of the weekly event study in the paper's Figure 1: daily treatment coefficients (θ_k) from the same trip-level Poisson QMLE on HPU intensity (log-mileage exposure offset; weekday, hour, and driver fixed effects), each a semi-elasticity relative to day -1. Standard errors are clustered at the driver level; bars are 95% confidence intervals. The red dashed line marks the intervention date; the grey dotted line marks the onset of the pandemic-declaration week (event day 28). A joint test of all 19 daily pre-period coefficients does not reject parallel pre-trends ($F = 1.06$, $p = 0.391$).

slightly larger $+0.0030$ hours ($p = 0.09$), which the controls absorb. Collapsing to driver-level pre/post changes gives a simple difference-in-differences of $+0.0019$ hours, but this is entirely right-tail-driven—it falls to zero (indeed marginally negative) after excluding the five treated drivers (of the 373 in this trip-level diagnostic panel) with the largest swings, and a randomization-inference test that permutes treatment across drivers (500 reps) returns a two-sided $p = 0.66$. The substitution hint present in earlier drafts does not survive the adopted sample window: among treated drivers, the correlation between the HPU change and the hours change is -0.07, and regressing the hours change on the HPU change gives a slope of -0.08 ($p = 0.17$), statistically indistinguishable from zero. A battery of coarser daily activity margins agrees: on a balanced driver$\times$day panel of the experiment sample (inactive days zero-filled; driver + day fixed effects, driver-clustered), daily trip counts, distinct active days, the probability of driving at all on a day, and total daily drive time all move by less than 1.4% of their pre-period means ($p \geq 0.69$ throughout). We therefore read the labor-supply margin as showing no robust response to the message; this is consistent with the HPU-intensity result, which is measured on the within-trip distraction margin and is not an artifact of a change in driving time. (Diagnostics: `55_experiment_hours.R`, run on the event-day < 34 sample.)

Table D8: Experiment: Average Treatment Effects and Passenger Targeting, All Outcomes

	Average effect (trip)				Passenger targeting (trip segment)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel: Dep. Var. = HPU Intensity								
Post × Treated	−0.2098**	−0.2095**	−0.2114***	−0.1956***	−0.1252*	−0.1242*	−0.1328***	−0.1224**
	(0.0967)	(0.0960)	(0.0542)	(0.0525)	(0.0697)	(0.0697)	(0.0471)	(0.0547)
Post × Treated × Passenger					−0.2000**	−0.1994**	−0.1865*	−0.1982*
					(0.0913)	(0.0911)	(0.0978)	(0.1044)
Panel: Dep. Var. = Harsh Acceleration								
Post × Treated	−0.0125	−0.0116	0.0814	0.1293	0.0497	0.0527	0.0914	0.1681*
	(0.0713)	(0.0718)	(0.0932)	(0.1006)	(0.0821)	(0.0821)	(0.0842)	(0.0906)
Post × Treated × Passenger					−0.0936	−0.0977	−0.0267	−0.0702
					(0.0701)	(0.0692)	(0.0738)	(0.0726)
Panel: Dep. Var. = Harsh Braking								
Post × Treated	0.0022	0.0051	0.0273	0.0365	−0.0140	−0.0108	0.0360	0.0417
	(0.0448)	(0.0452)	(0.0350)	(0.0341)	(0.0513)	(0.0516)	(0.0453)	(0.0446)
Post × Treated × Passenger					0.0250	0.0242	−0.0140	−0.0112
					(0.0419)	(0.0414)	(0.0435)	(0.0481)
Panel: Dep. Var. = Harsh Cornering								
Post × Treated	0.0245	0.0254	0.0166	−0.0277	−0.0130	−0.0127	−0.0369	−0.0683
	(0.0735)	(0.0728)	(0.0580)	(0.0697)	(0.0754)	(0.0753)	(0.0672)	(0.0827)
Post × Treated × Passenger					0.0903	0.0928	0.1185	0.0996
					(0.1045)	(0.1041)	(0.1078)	(0.1131)
Panel: Dep. Var. = Hours Driven								
Post × Treated	0.0035	0.0019	0.0010	0.0002	0.0021	0.0012	0.0007	0.0004
	(0.0030)	(0.0025)	(0.0021)	(0.0017)	(0.0014)	(0.0013)	(0.0015)	(0.0016)
Post × Treated × Passenger					−0.0007	−0.0004	−0.0003	−0.0009
					(0.0027)	(0.0027)	(0.0027)	(0.0026)
Observations	464,045 trips				849,488 segments			
Trip-segment controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Driver + Date + Hour FE	No	No	Yes	Yes	No	No	Yes	Yes
Location-pair FE	No	No	No	Yes	No	No	No	Yes

Notes: $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$. Driver-clustered SEs. Every column reproduces the corresponding specification of Table 5 for each outcome: columns (1)–(4) are the TRIP-level average effect (adding controls, then driver+date+hour fixed effects, then geohash-5 location-pair fixed effects); columns (5)–(8) add the Post×Treated×Passenger interaction at the TRIP-SEGMENT level over the same progression. HPU intensity is Poisson QMLE on handheld-phone miles with a log-mileage offset at the trip level and on handheld-phone seconds with a log-driving-time offset at the segment level, exactly as in Table 5; the harsh-event counts are Poisson QMLE with a log-distance offset; hours driven is linear.

E Data and Algorithm: From Sensors to Handheld Phone Use Behavior ("HPU")

Our behavioral data is constructed from raw telematics data in three stages. First, we collect off the smartphone 25Hz three-dimensional streams of accelerations, rotations, and force of gravity based on the phone's accelerometer and gyroscope.[1] We also collect 1Hz streams of GPS location records. These raw sensor records are serialized using Google's Protocol

[1]An accelerometer is a compact device designed to measure non-gravitational acceleration. A gyroscope measures the phone's rotation rate about each of its axes.

Table D9: Robustness: Effect on the Extensive Margin of HPU

	Average effect (trip)				Passenger targeting (trip segment)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post × Treated	−0.0256	−0.0258	−0.0301**	−0.0292**	−0.0254*	−0.0257*	−0.0304**	−0.0283**
	(0.0160)	(0.0159)	(0.0134)	(0.0133)	(0.0151)	(0.0150)	(0.0124)	(0.0119)
× Passenger					−0.0084	−0.0081	−0.0067	−0.0070
					(0.0108)	(0.0108)	(0.0108)	(0.0107)
Trip-segment controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Driver + Date + Hour FE	No	No	Yes	Yes	No	No	Yes	Yes
Location-pair FE	No	No	No	Yes	No	No	No	Yes
Observations	464,045 trips				849,488 segments			

Notes: $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$. Driver-clustered SEs. Dependent variable: $\mathbf{1}\{\text{HPU} > 0\}$ (extensive margin; linear probability). A trip is the aggregation of its trip segments (pickup-driving P2 and passenger-aboard P3). Columns (1)–(4) are the TRIP-level average effect, adding controls, then driver+date+hour fixed effects (column 3, the canonical specification), then geohash-5 location-pair fixed effects (column 4, matching the precipitation analysis). Columns (5)–(8) are at the TRIP-SEGMENT level, with the outcome an indicator of any handheld use in the segment (linear probability), and add the Post×Treated×Passenger interaction (first row: the pickup-segment (P2) effect; ×Passenger row: the passenger-segment (P3) differential), mirroring the same fixed-effects progression. Linear probability model for any handheld phone use in the trip/segment.

Table D10: Robustness: Effect on HPU Intensity (linear)

	Average effect (trip)				Passenger targeting (trip segment)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post × Treated	−9.1761**	−9.1656**	−9.2018***	−9.2894***	−0.0054	−0.0053	−0.0057**	−0.0056**
	(4.2583)	(4.2619)	(2.5519)	(2.5032)	(0.0034)	(0.0034)	(0.0024)	(0.0023)
× Passenger					0.0016	0.0016	0.0017	0.0017
					(0.0025)	(0.0025)	(0.0025)	(0.0025)
Trip-segment controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Driver + Date + Hour FE	No	No	Yes	Yes	No	No	Yes	Yes
Location-pair FE	No	No	No	Yes	No	No	No	Yes
Observations	464,045 trips				849,488 segments			

Notes: $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$. Driver-clustered SEs. Dependent variable: handheld-phone miles per mile (×1000; linear). A trip is the aggregation of its trip segments (pickup-driving P2 and passenger-aboard P3). Columns (1)–(4) are the TRIP-level average effect, adding controls, then driver+date+hour fixed effects (column 3, the canonical specification), then geohash-5 location-pair fixed effects (column 4, matching the precipitation analysis). Columns (5)–(8) are at the TRIP-SEGMENT level, with the outcome the segment's handheld share of driving time (linear), and add the Post×Treated×Passenger interaction (first row: the pickup-segment (P2) effect; ×Passenger row: the passenger-segment (P3) differential), mirroring the same fixed-effects progression. Linear model on handheld-phone miles per mile (×1000).

Buffer, compressed down to a very small data package over a small time period, and transmitted to our data center in real-time together with other preexisting data collected for navigation and for normal app functioning.

In the second stage, to process these raw signals, we use a hybrid heuristic- and ML-based approach. We first use a standard low-pass filter to attenuate excessive noise and further regularize the sensor data with a gravity removal procedure outlined in Hassan

Table D11: Experiment: Full-Control Difference-in-Differences vs. Strict Randomized Control

	(1) Full untreated control (DiD)	(2) Strict randomized control
Post × Treated (intensity)	-0.2114^{***}	-0.2062^{***}
	(0.0542)	(0.0630)
Post × Treated (extensive)	-0.0301^{**}	-0.0410^{***}
	(0.0134)	(0.0146)
Pre-trend F $[p]$	1.06 [0.391]	0.78 [0.727]
Treated / control drivers	343 / 2,160	343 / 407
Trips	464,045	138,200

Notes: $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$. Driver-clustered SEs. Both columns are the canonical specification of Table 5 column (3) — controls and driver, date, and hour fixed effects — for the intensity margin (Poisson QMLE on handheld-phone miles per mile, log-mileage offset; semi-elasticity) and the extensive margin (linear probability on any handheld use). Column (1) uses the full set of untreated targeted drivers as the comparison, the difference-in-differences of Table 5, identified by parallel trends. Column (2) restricts the comparison to the within-window randomized control group — drivers who logged on during the intervention window and were randomized out of treatment — the clean experimental contrast. The pre-trend row is the joint Wald F (p-value in brackets) on the pre-period event-day leads.

Table D12: Pre-Treatment Balance: Treated vs. Randomized and Matched Controls

	Treated	Randomized control		Matched control	
Variable	($N = 340$)	Mean ($N = 398$)	Diff	Mean ($N = 2134$)	Diff
HPU frequency (pre)	0.0217	0.0291	-0.0075**	0.0277	-0.0060***
Harsh accel. rate (pre)	0.0003	0.0003	-0.0000	0.0003	-0.0000
Harsh brake rate (pre)	0.0006	0.0007	-0.0001*	0.0006	-0.0001*
Avg. drive duration (sec)	696.2950	701.6741	-5.3792	706.4923	-10.1973
Avg. distance (mi)	8.5894	8.7382	-0.1488	8.7962	-0.2068
Num. trips (pre-period)	126.3676	133.0226	-6.6550	128.4897	-2.1220
Age (years)	38.4930	38.0308	0.4622	38.9313	-0.4384
Tenure (years)	1.4720	1.6937	-0.2216**	1.6311	-0.1590**
Past accidents	0.0088	0.0126	-0.0037	0.0122	-0.0034
Past unsafe rides	0.4647	0.5930	-0.1283	0.5712	-0.1065
Past safety tickets	0.4647	0.6080	-0.1433*	0.5515	-0.0868
Past rides completed	262.6000	265.8216	-3.2216	260.4958	2.1042
Past total earnings ($)	2530.5098	2626.7727	-96.2629	2562.5352	-32.0254
Female (proportion)	0.1559	0.1784	-0.0225	0.1860	-0.0302

Notes: $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$ (two-sample t-test of the treated–control difference). Pre-treatment driver-level means over days -1 and earlier. Counts fall slightly below the design counts (343 treated, 407 randomized, 2,160 matched) because balance requires at least one pre-period driving day; 3 treated, 9 randomized, and 26 matched drivers have none. *Treated*: messaged top-HPU drivers. *Randomized control*: drivers who logged on in the 30-minute experimental window and were randomized to control. *Matched control*: the wider matched pool used in the difference-in-differences regressions (the randomized controls are a subset).

et al. (2018). We then orient the phone with respect to the vehicle based on GPS movement patterns, which reveal the resting position of the phone within the vehicle (Skog and Handel 2009). These steps make telematics sensor data interpretable from both the phone's and the vehicle's perspectives—the acceleration and gyration of the phone, adjusted for its speed, pin down the orientation of the vehicle relative to the phone.

In the third stage, we construct sensor data features and identify behaviors using an in-house machine-learning algorithm. We first identify candidate "events," which are defined as abnormal movements in at least one of the sensor data streams. Operationally, we aggregate each sensor data stream over small time intervals and calculate moments of the data. We then detect events using pre-specified moment thresholds. The combination of detected events over several data streams identifies behaviors such as HPU.

Many phone movements during a trip are not handheld. Drivers often keep their phones in cup-holders or on dashboards; disruptive vehicle movements such as going through speed bumps or collision events themselves can also cause large sensor-data fluctuations. To isolate genuine handheld interaction, the algorithm requires the joint occurrence of (i) phone motion independent of the vehicle's motion, (ii) nonzero driving speed, and (iii) sustained anomalous motion over a nontrivial duration (Engelbrecht et al. 2015; Wahlström et al. 2017).

Nonetheless, compared to other risky driving behaviors such as speeding, HPU is particularly easy to identify because its detection is almost entirely unaffected by environmental factors such as speed limits and weather events. Intuitively, HPU identification primarily relies on abnormal gyration of the smartphone itself: events associated with abnormally large variance in gyroscope sensors tend to indicate either HPU or harsh turns, but HPU is associated with shorter durations of gyroscope-sensor abnormality, which results in higher kurtosis, as well as lower variance in speed. Our training and validation datasets are proprietary and labeled based on test drives and video footage of test drivers, supplemented with independent labels from FleetCam. This allows accurate and synchronized labeling of HPU and the corresponding sensor movements. The resulting algorithm achieves over 99% precision and over 75% recall.[2]

[2]The focus on precision is to limit liabilities associated with intervention based on false-positive identification of HPU. We briefly discuss the implications of our results for optimal precision-recall choice in the conclusion.

References

Engelbrecht, Jarret, Marthinus J. Booysen, Gert-Jan Van Rooyen, and F. J. Bruwer (2015). "Survey of smartphone-based sensing in vehicles for intelligent transportation system applications". In: *IET Intelligent Transport Systems* 9.10, pp. 924–935.

Hassan, Mohammed Mehedi, Md Zia Uddin, Amr Mohamed, and Ahmad Almogren (2018). "A robust human activity recognition system using smartphone sensors and deep learning". In: *Future Generation Computer Systems* 81, pp. 307–313.

Prentice, Ross L and Ronald Pyke (1979). "Logistic disease incidence models and case-control studies". In: *Biometrika* 66.3, pp. 403–411.

Skog, Isaac and Peter Handel (2009). "In-car positioning and navigation technologies—A survey". In: *IEEE Transactions on Intelligent Transportation Systems* 10.1, pp. 4–21.

Wahlström, Johan, Isaac Skog, and Peter Händel (2017). "Smartphone-based vehicle telematics: A ten-year anniversary". In: *IEEE Transactions on Intelligent Transportation Systems* 18.10, pp. 2802–2825.