Safer Driving for a Price: Evidence on Behavior and Habits in Kenyan Minibuses

David Schönholzer ⓡ Gregory Lane ⓡ Erin Kelley*

June 20, 2026

Abstract

Road traffic injuries are the leading cause of death among people aged 5–29. Where traffic laws are weakly enforced, it is unclear whether road safety interventions can change behavior. We run a randomized controlled trial among 203 Kenyan minibuses, testing whether an incentive scheme improves safety overall or merely shifts unsafe driving to unmonitored times. Drivers reduce speeding by 29 percent and harsh braking by 13 percent, improving a safety index by 0.096 standard deviations. Labor-supply and earnings effects are modest, with little displacement across times or places. Improvements fade quickly, suggesting lasting change requires sustained enforcement or stronger incentives.

Keywords: Road Safety, Informal Transit, Enforcement

JEL Classification: O12, O18, K42

*ⓡ Author order randomized. Kelley: Harris School of Public Policy, University of Chicago (email). Lane: Harris School of Public Policy, University of Chicago (email). Schönholzer: Department of Economics, University of Southern California (email). We thank Vittorio Bassi for useful comments. We are extremely grateful to the entire team at EchoMobile (Jeremy Gordon, Boris Maguire, James Mwangi, Vitalis Omutanyi, James Langat, James "Wambururu" Kariuki, and Kevin Onyango) for exceptional field management and implementation, as well as Peter Waiganjo Wagacha for guidance and support. We acknowledge financial support from DIL, IGC, and FCDO. IRB approvals have been obtained at UC Berkeley (IRB 2015-02-7180) and the University of Nairobi (KNH-ERC/A/417). PAP is at the AEA registry under AEARCTR-0001482. All errors are our own.

1 Introduction

Road traffic injuries are the leading cause of death among children and young adults aged 5–29, and they occur disproportionately in low-income countries. According to the WHO, low- and middle-income countries account for 60% of the world's vehicles but 92% of road traffic fatalities (WHO, Geneva, 2020). Many countries respond with strict regulations: speed limits, penalties for reckless driving, and removing unsafe vehicles and drivers from the road. But the effectiveness of these measures depends on enforcement capacity, which is often limited in low-income countries. Thus, a central policy question is whether raising the cost of unsafe driving can improve safety even when enforcement capacity is weak.

This paper tests that question in Nairobi, Kenya, where traffic regulations are weakly enforced. We introduce a monetary incentive scheme that prices unsafe driving directly, a mechanism insurers, fleet owners, the police, or transport authorities could plausibly run. We focus on the informal public transit industry, which is widely recognized as contributing to these unsafe road conditions. In Kenya, road crashes frequently involve minibuses ("matatus"), which account for a disproportionately large share of passenger casualties relative to their share of the vehicle fleet (Macharia et al., 2009).

Specifically, we randomly assign 203 minibus drivers to a treatment group (77 drivers) and a control group (126 drivers). Treated drivers receive a daily bonus from which we deduct penalties for unsafe driving for one to two months.[1] The mechanism works as follows. A GPS device installed in every study vehicle records alerts for speeding, sharp braking, over-acceleration, and sharp turning wherever and whenever the matatu is in service, and drivers are aware that their driving is monitored. Over an initial calibration period of 20 driving days, we use these alerts to construct each driver's baseline distribution of alerts. Once treatment begins, drivers start each day with a 600 KES bonus (about 6 USD at the time, 60% of average daily earnings), and we deduct from it in increasing amounts as the day's unsafe driving exceeds the driver's baseline average. Each day's driving is scored once, around 10 p.m., and drivers receive an SMS that evening notifying them of the amount they lost because of unsafe driving maneuvers. Evaluating drivers against their own baseline, rather than against one another, avoids penalizing inherently less safe drivers simply for being less safe than their peers. Because the penalty is deducted from a bonus we provide, no driver is made worse off in expectation by the intervention. If drivers respond less strongly to losing part of a bonus than to losing their

[1]This experiment was implemented within the same fleet of vehicles in which owners were independently and randomly assigned to receive information about their drivers' behavior (analyzed in Kelley et al. 2024).

own money, our estimates can be interpreted as a lower bound.[2]

The intervention is designed to raise the price of unsafe driving. While the incentive is active, we expect treated drivers to reduce risky behavior. However, the intervention's value depends on two other phenomena—displacement and persistence—not just on whether the direct effect exists. We define displacement as follows. Unsafe driving is a means to raise earnings: drivers speed and maneuver aggressively to pick up more passengers. If safer driving reduces earnings, drivers may compensate by shifting risky behavior to periods when enforcement is weaker. In our setting, because the day's score is finalized at 10 p.m., drivers could in principle shift unsafe driving behavior to after that time in order to avoid being penalized. This is one of the displacement channels we examine below. Alternatively, drivers could take more off-the-books jobs outside the metro area rather than driving more safely overall. This pattern of displacement is documented in other enforcement settings: fixed speed cameras produce localized slowdowns followed by re-acceleration once drivers are past the camera, while average-speed cameras (which measure speed over a stretch of road rather than at a single point) work better precisely because they leave fewer opportunities to shift risk across space and time (De Pauw et al., 2014; Høye, 2014). In incentive environments more broadly, partial monitoring tends to shift effort toward incentivized tasks without improving overall performance (Lazear, 2006; Reback, 2008). To measure this, we use granular data from the monitoring device to analyze how driving behavior shifts throughout the day.

We also consider persistence, after the incentive ends. If driving safely becomes habit —through repeated exposure lowering the driver's customary level of risk and making a return to old behavior costlier (Becker and Murphy, 1988; Hussam et al., 2022)—the safety gains may outlast the contract. If not, behavior reverts as soon as pricing stops. To measure whether longer contract exposure builds habit, we randomly assign treated drivers to one or two months of incentives and compare persistence across the two exposure lengths.

We document three main results. First, we demonstrate the impact of cash incentives on driver behavior, starting with driver safety. We find that the intervention increases a composite safety index by 0.096 standard deviations, driven by reductions in speeding

[2]Two forces suggest that an out-of-pocket fine would be more effective than the bonus scheme we study, implying that our experiment may estimate a lower bound on the effects of monetary penalties. First, under concave utility, the marginal value of income declines as earnings increase. As a result, losing one shilling from a bonus is less painful than paying an equivalent fine out of pocket. Second, a fine directly reduces take-home pay, pushing more drivers below their daily earnings target. If the fine is large enough that unsafe driving becomes unprofitable on net, drivers who are below their target may respond by driving more carefully to protect their remaining earnings, rather than driving more aggressively to make up the shortfall.

(0.297 alerts per day, or 29%) and sharp braking (0.158 alerts per day, or 13%). Next, we look at whether drivers adjust any other margins of behavior. We find a moderate increase in labor supply and no meaningful effects on the other margins drivers can adjust (kilometers driven, revenue, or take-home pay), so drivers achieve these safety gains by working somewhat longer while keeping their earnings intact.

Second, we test for strategic responses to enforcement via displacement. In particular, we investigate whether the absence of earnings effects masks any intra-day displacement: drivers may drive more safely and earn less during the day when the safety score is being computed, and then drive recklessly at night once the incentives are locked in, to make up for the lost revenue; or they may seek out weaker-enforced areas by taking more off-the-books jobs outside the metro area. We find little evidence of displacement, so the safety gains during active treatment appear genuine and are achieved with only a moderate increase in labor supply that leaves earnings intact.

Third, we investigate the role of persistence and habit formation. We show that safety gains fade once incentives end, though the decline is somewhat slower for drivers randomly assigned to longer exposure. This pattern is consistent with limited habit formation and suggests that sustained incentives may be needed to produce durable improvements.

Together, these results suggest that directly paying drivers based on safety performance is a viable policy in settings with generally weak enforcement and notoriously risky driving. Given the widespread availability of tracking devices in transit vehicles in low-income contexts (usually for safety and insurance reasons), this policy could be adopted by authorities or insurance companies. Under conservative assumptions about lives saved and damage prevented, our estimates suggest that the intervention was cost-effective.

This study makes three primary contributions. First, we contribute to the literature on improving safety in informal transit. Related work in the Kenyan minibus industry has studied the impact of different safety interventions: passenger pressure on drivers (Habyarimana and Jack, 2011), driver compensation structures (Johnson et al., 2015), and owner monitoring (Kelley et al., 2024). We study a different lever: tying the driver's daily pay directly to a tracker-based measure of how safely they drove that day. The approach is straightforward to implement: it relies only on an in-vehicle tracker to calibrate and pay out the incentive, mirroring telematics-based contracts used by insurers in high-income countries. A rough back-of-the-envelope calculation suggests the intervention need only avert a small number of statistical deaths over the study window to cover its net transfer cost. More broadly, the approach of paying firms to mitigate negative externalities has found positive impacts (Brown et al., 2024). However, the gains we find fade once the incentive ends, which suggests durable improvement requires keeping the pay-for-safety arrangement in place rather than running it as a temporary campaign.[3]

The paper also contributes to the literature on strategic responses to monitoring, specifically, whether agents substitute risky behavior to times when monitoring is weaker. Zou (2021) shows that intermittent pollution monitoring in the United States leads to worse air quality on unmonitored days, as cities suppress emissions when monitors are active and let them rise when monitors are off. There is reason to expect drivers in our setting may respond similarly. The taxi labor-supply literature finds that drivers often work toward a daily earnings target (Camerer et al., 1997; Chou, 2002; Farber, 2008; Crawford and Meng, 2011; Ashenfelter et al., 2010; Farber, 2015), so if safer driving cut into daytime revenue, drivers could compensate by working longer or by driving more unsafely in unmonitored hours. Our experiment lets us test for this: we observe whether drivers adjust how many hours they work *and* how carefully they drive when the cost of unsafe driving rises. We find some compensation on the hours margin but not through displacement: drivers work moderately longer hours, and the 30-second tracker data show no shift of unsafe driving to lower-enforcement hours. Importantly, driver revenue is essentially unchanged, consistent with drivers working somewhat longer to protect their earnings. This pattern is consistent with the bicycle-taxi setting of Dupas et al. (2020), where Kenyan transit workers work toward a daily earnings target: if drivers hit their target despite driving more safely, little compensation on other margins is needed.

Finally, the paper contributes to the empirical literature on habit formation (Becker and Murphy, 1988). Field experiments have tested whether temporary incentives create durable behavior change in exercise (Charness and Gneezy, 2009; Royer et al., 2015), handwashing (Hussam et al., 2022), and labor supply (Cefala et al., 2024), often finding lasting effects. Safe driving is a distinctive testing ground: unlike health behaviors, whose benefits accrue mostly to the incentivized individual, safer driving competes directly with the driver's earnings and benefits mostly others. Our results suggest some habit formation, but not enough to sustain gains: longer exposure to incentives produces slightly more persistence, but safety improvements fade once incentives end.

The rest of the paper is organized as follows. Section 2 describes the Nairobi setting, Section 3 develops a simple model, Section 4 details the experiment and empirical framework, and Section 5 presents the safety, labor-supply, and persistence results, and Section

[3]Recent telematics trials confirm that such tools can reduce risky driving while active (Stevenson et al., 2021; Malekpour et al., 2023), but it is unclear whether improved driving habits form due to such interventions.

6 concludes.

2 Setting: Informal Transit in Nairobi

Nairobi's transportation system emerged after Kenya's independence in 1963 (Mutongi, 2017). Small-scale entrepreneurs responded to growing demand for mobility by retrofitting old vehicles and ferrying passengers between the suburbs and the urban center. These buses became known as "matatus," from the Kikuyu word for three, in reference to the early ticket price of thirty cents, paid as three ten-cent coins. The businesses were legalized in 1973 but remained largely unregulated for the next three decades. Regulation has expanded on paper but unevenly in practice. In 2003, the government passed the Michuki rules, requiring buses to install speed limiters and safety belts and to ensure all drivers held valid licenses (Michuki, 2003). In practice, however, these regulations are unevenly enforced. In 2010, the Ministry of Transport issued a further directive requiring all minibus owners to join a Savings and Credit Cooperative (SACCO) licensed to a specific route (McCormick et al., 2013). SACCOs leave daily management of the vehicle to the owner but coordinate scheduling along the designated route, resolve disputes between owners, ensure compliance with National Transport and Safety Authority (NTSA) regulations, and provide financial services to owners and drivers (Rasmussen, 2012; Cervero and Golub, 2007; Behrens et al., 2015; McCormick et al., 2013; Williams et al., 2015; Mutongi, 2006, 2017). More recently, there have been efforts to regulate fares and limit the number of buses operating in the market. However, it is unclear whether these regulations have been enforced at all (Mwende, 2025).

Matatus remain Nairobi's only dependable transit system to this day. Rough estimates suggest 15,000 to 20,000 buses circulate through the city, swerving on and off the road to collect passengers along designated routes. The industry remains almost entirely locally owned: private entrepreneurs increasingly purchase 33-seat minibuses rather than 14-seat ones, join a SACCO, and hire a driver to operate the vehicle along the SACCO's route. Matatu owners often invest in the comfort and appearance of their vehicles: colorful interior and exterior designs, and onboard amenities such as televisions or wifi (Reed, 2018). The more attractive vehicles can charge up to twice the price of standard ones. Fares range from roughly 0.5 to 1.5 USD within the city center and 1 to 5 USD for trips to the outskirts.

Matatu owners have settled on a fixed rent contract with limited liability that is negotiated daily (Kelley et al., 2024). Owners rent their vehicles to a driver every morning for an agreed upon "target" price. Unlike the taxi systems in many high-income countries, the driver is expected to deliver this amount at the end of the day once all the fares have been collected. This is primarily because drivers have limited capital and cannot afford to pay the amount up front. Drivers are the residual claimants in this contract and keep everything they earn above the target. The owner is not allowed to revise the terms of the contract and claim more revenue if the driver has had a good day.

Drivers operate under demanding conditions. A typical shift can run twelve to sixteen hours, navigating heavy traffic, unpredictable road conditions, with limited rest opportunities. Earning the target on any given day is far from guaranteed: passenger volumes fluctuate with weather, road closures, fuel prices, and competition from other vehicles on the same route. In the event that the driver cannot make the target, they are supposed to provide the total revenue they earned to the owner. In practice, drivers under-report total revenue to make sure they have some income left over. If they fail to make the target too many times, or they are caught under-reporting too frequently, they will be fired. Our data suggests that owners and drivers work with each other for two years on average (Kelley et al., 2024).

Because every additional passenger raises the driver's take-home pay while the costs of wear and tear on the vehicle are borne by the owner, drivers have strong incentives to pick up as many passengers as possible, even when doing so requires aggressive or unsafe driving (Raynor and Mirzoev, 2014; Kelley et al., 2024). The consequences for road safety are stark: between 3,000 and 13,000 people die annually in traffic incidents in Kenya, and at least 30% of these cases involve matatus (WHO, 2015). Studies of Kenyan road injuries point to speeding, abrupt maneuvers, weak enforcement, and corruption in traffic policing as central contributors (Macharia et al., 2009; Raynor and Mirzoev, 2014). Our prior work (Kelley et al., 2024) shows that providing owners with information about unsafe driving through vehicle monitoring devices does not lead to safer driving. Owners care primarily about *vehicle damage*, not passenger safety, and the two are not always aligned.

3 A Simple Model of Imperfect Enforcement under Behavioral Labor Supply

The experiment temporarily raises the expected price of unsafe driving. The direct effect should push drivers toward safer behavior, and repeated exposure may build habits that persist after the incentives end. But drivers may also respond by finding other ways to hit their earnings target. They may work longer hours, or drive aggressively at times or in places where pricing is absent, potentially offsetting any safety gains. The taxi labor-supply literature documents that drivers tend to behave as if working toward a daily earnings target (Camerer et al., 1997; Crawford and Meng, 2011; Farber, 2015), and matatu drivers may be no different. We use a simple model to organize these competing effects.

Let h_t denote hours worked on day t, s_t unsafe driving intensity, and k_t the driver's habitual level of risky driving, a stock formed from his own past behavior (Becker and Murphy, 1988) as described below. Daily fare earnings are $p(h_t + \sigma s_t)$, where p is the fare return to effective service and σ captures how easily risky driving substitutes for longer hours in generating earnings, by shortening trips or increasing seat turnover. Let T denote a daily earnings target (inclusive of the fixed remittance drivers owe owners) and $\lambda \geq 0$ the strength of target pressure when earnings fall short of that target. Let $\pi_t \in [0, 1]$ denote enforcement intensity on day t and $\tau > 0$ denote the penalty or reward-equivalent price per unit of risky driving, so the expected price of unsafe driving is $\pi_t \tau$. The experiment is a temporary increase in $\pi_t \tau$: pricing becomes salient while incentives are active, then returns to zero when they end. The driver's one-day problem is

$$\max_{h_t, s_t} \underbrace{p(h_t + \sigma s_t)}_{\text{fare earnings}} - \underbrace{c(h_t)}_{\text{cost of hours}} - \underbrace{d(s_t, k_t)}_{\text{habit cost of risk}} - \underbrace{\pi_t \tau s_t}_{\text{expected penalties}} - \underbrace{\lambda \max\{T - p(h_t + \sigma s_t), 0\}}_{\text{target pressure}},$$

where $c(h_t)$ is the convex disutility of hours, $d(s_t, k_t)$ is the private cost of unsafe driving (increasing and convex in s_t, with marginal cost decreasing in habitual risk, $d_{sk} < 0$: driving as riskily as one is used to feels cheap), and enforcement raises the private cost of unsafe driving during priced periods.

Whether temporary pricing causes drivers to develop safe driving habits that persist even when enforcement is relaxed depends on how the habit evolves. Following Becker and Murphy (1988), the habit stock is an average of the driver's own past behavior:

$$k_{t+1} = (1 - \rho) k_t + \rho\, s_t,$$

where $\rho \in (0, 1)$ governs how quickly the habit adjusts to recent behavior.[4] Habit formation is the complementarity between past and current behavior (Becker and Murphy, 1988; Hussam et al., 2022). A spell of safe driving lowers k_t, which raises the marginal cost of reverting to risk and so sustains safer driving after pricing ends. Whether this feedback fades or becomes self-sustaining is an empirical question: if the complementarity is weak, behavior reverts quickly once pricing ends; if it is strong enough, a temporary

[4]We assume drivers are myopic with respect to this stock: each day they take k_t as given and do not internalize the effect of today's driving on future habit. The habit thus operates as a reduced-form adaptive process rather than the forward-looking accumulation of Becker and Murphy (1988).

incentive can in principle tip drivers into a permanently safer habit. Existing transport-safety evidence points to temporary after-effects from enforcement, but not obviously to permanent behavioral change (Hauer et al., 1982; Vaa, 1997; Inada et al., 2022).

When earnings fall short of the target, optimal interior behavior satisfies

$$\underbrace{c'(h_t)}_{\text{marginal cost of hours}} = \underbrace{(1+\lambda)\,p}_{\text{earnings value of hours}},$$

$$\underbrace{d_s(s_t,k_t)}_{\text{habit cost of risk}} + \underbrace{\pi_t\tau}_{\text{enforcement price}} = \underbrace{(1+\lambda)\,p\,\sigma}_{\text{earnings value of risk}}.$$

When earnings exceed the target, the same conditions hold with $\lambda = 0$. These conditions deliver three predictions that the experiment is designed to test:

1. **Direct effect.** While incentives are active, a higher price of risky driving $\pi_t\tau$ lowers s_t via the habit cost of risk, $d_s(s_t, k_t)$. In the second condition, $\pi_t\tau$ adds directly to the marginal cost of risk, so s_t must fall to restore equality.

2. **Displacement.** If earnings targets bind (large λ) or risky driving readily substitutes for hours (large σ), safer driving is offset by longer hours or by shifting risk to periods with weaker enforcement. Target pressure enters both conditions through the multiplier $(1+\lambda)$, which raises the earnings value of hours and of risk alike: pushing up h_t and rewarding risk wherever the enforcement price is low.

3. **Persistence.** After incentives end, safe driving persists only if the habit has shifted: once $\pi_t\tau$ returns to zero, only a lower k_t keeps the marginal cost of risk elevated ($d_{sk} < 0$). If the habit has shifted, safe driving fades only gradually once incentives end, and it fades more slowly in the two-month arm, whose habit stock has fallen further.

Having specified how an enforcement policy can affect the behavior of matatu drivers, we now describe the experimental design to test these three predictions.

4 Experimental Design

Sample Selection. Our sample of drivers is recruited through a separate experiment involving minibus owners, described in Kelley et al. (2024). As part of that study, we conducted an extensive recruitment effort in late 2015, contacting 61 SACCOs operating along 9 routes across the city, as shown on the map in Appendix Figure A1. We organized meetings with matatu owners within each SACCO and informed them that we had developed a tracking device to be installed in all participating vehicles. Owners were told that a randomly selected subset would immediately receive information generated by the device (the *information treatment*), while the remainder would receive access after a six-month delay. We successfully recruited 255 minibus owners and installed tracking devices in their vehicles. During installation, a field manager met separately with the driver to explain how the device worked. These drivers constitute our study sample. All drivers know that a tracking device is installed in the vehicle.

Tracker Device. The research team created a new monitoring system for minibuses that was considerably cheaper, more flexible and more powerful than traditional tracking devices. The physical tracking units were procured for 125 USD from a company in the United States (CalAmp). They feature GPS, internal back-up battery packs, 3-axis accelerometer for motion sense, tilt and impact detection. The device was designed to capture and transmit the 95th percentile and average forward/backward/lateral/vertical acceleration, as well as the 95th percentile and average forward/backward jerk. The device was also calibrated to generate alerts for every instance of vehicle speeding, over-acceleration, sharp braking and sharp turning. These safety alerts were calculated by an internal algorithm built into the CalAmp device with threshold parameters as inputs. Further processing of the CalAmp system data on the server provided additional measures of interest including the total number of kilometers traveled that day, the total time the matatu was running, and a safety index (from aggregating the day's safety alerts).

Randomization. Four months after the *information treatment* from Kelley et al. (2024) was implemented, we launched the *safety treatment* and further randomized minibus drivers into two groups: a treatment group (77 drivers), who received cash incentives tied to safe driving behavior, and a control group (126 drivers), who did not.[5] Drivers in the treatment group were then randomly assigned to either a one-month (41 drivers) or two-month (36 drivers) incentive period. This design allows us to examine (1) whether incentives affect safety, and (2) whether any improvements persist after the incentives are removed.

The incentive amount each treated driver receives is based on a daily safety score, calculated as follows. During a 20-day calibration period (excluding days with less than

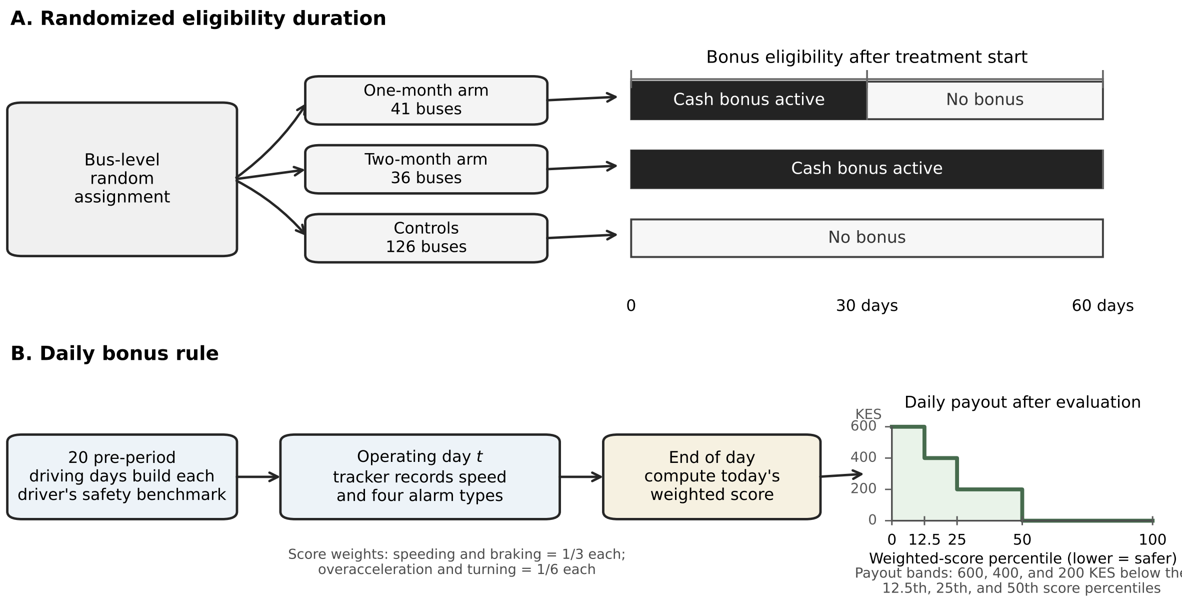
[5]The safety treatment partly overlapped in calendar time with the information experiment of Kelley et al. (2024), with which it was independently randomized. For a subset of treated drivers, the cash incentive was administered in two separate one-to-two-month stints; the second began after these drivers' participation in the information experiment had ended, providing a treatment window outside the overlap. Our main estimates are robust to using only this later, non-overlapping stint (Appendix Table A5).

30 km driven), we record (1) the number of speeding, hard braking, sharp turning, and overacceleration alerts, and (2) the number of hours worked each day. For each driver, day, and alert type, we compute an alert rate by dividing the number of alerts by hours worked. We then construct a driver-specific distribution of these alert rates for each of the four alert types over the calibration period.

Once treatment begins, we score each new day by asking where the driver's alert rate falls in their own pre-period distribution, expressed as a percentile (lower percentiles indicate safer days). We combine the four percentiles into a single weighted average, with weights of 1/3 each for speeding and braking and 1/6 each for overacceleration and sharp turning. The daily payout follows directly from the composite percentile. The maximum daily bonus is 600 KES, roughly 60 percent of mean daily driver pay in the sample. Drivers receive the full 600 KES if their composite percentile is below 12.5 (driving as safely as on their safest 12.5% of pre-period days), 400 KES between 12.5 and 25, 200 KES between 25 and 50, and nothing above 50. Evaluating drivers against their own baseline rather than against one another is a deliberate design choice. It avoids penalizing inherently less safe drivers simply for being less safe than their peers, and ensures the incentive rewards improvement on each driver's own margin.[6] Figure 1 summarizes the experiment: Panel A shows the randomized eligibility durations and Panel B the daily scoring and payout rule.

[6]To reinforce the salience of the incentive, field officers followed up with drivers by phone, and in person during onboarding, to confirm they understood the daily SMS and how their driving translated into the bonus.

Figure 1: How the driver incentive worked



Notes: Schematic summary of the driver cash treatment. Panel A shows eligibility duration by arm; the treatment roster identifies 41 later-wave buses assigned to one month of eligibility and 36 assigned to two months. Panel B shows the day-level scoring logic and the stepped daily payout rule: each day's weighted alert score (weights of 1/3 on speeding and sharp braking, 1/6 on overacceleration and sharp turning) is evaluated against the driver's own benchmark distribution, and the payout is 600 KES below the 12.5th percentile of the weighted score (lower percentiles indicate safer days), 400 KES between the 12.5th and 25th, 200 KES between the 25th and 50th, and zero above the 50th. The maximum daily bonus of 600 KES is roughly 60 percent of mean daily driver pay in the analysis sample.

Data and outcomes. Our data come directly from the tracking devices installed in study vehicles, and we use them in two forms. The first is a daily panel of 41,728 positive-mileage bus-days, of which our main regressions use 37,353 (this excludes post-treatment observations, defined as days after incentives ended for both treatment groups). The second is the raw 30-second stream, which records location, forward and lateral acceleration, vertical acceleration, and jerk, which we use for within-day analyses to test for displacement of unsafe driving to later hours.

Appendix Table A1 shows that treatment and control drivers are balanced on baseline characteristics, as well as pre-treatment operational and safety outcomes. For treated buses, pre-treatment measures are calculated using up to 30 observed positive-mileage bus-days immediately preceding the start of treatment. For control buses, we assign placebo treatment start dates by drawing from the distribution of treated start dates within route and applying the same 30-observed-day window.

Our primary outcome is the composite safety index specified in our pre-analysis plan.

For each alert type, we divide the day's alert count by the average daily count on the vehicle's route over its first two weeks of data.[7] We then combine the four route-normalized rates into a weighted average, with weights of 1/3 on speeding and sharp braking and 1/6 on over-acceleration and sharp turning, flip the sign on the sum so that higher values indicate safer driving, and standardize using the control-group mean and standard deviation. The index thus tracks the same four behaviors, with the same weights, as the bonus rule, but benchmarks each day against route averages rather than against the driver's own calibration distribution used to set payouts. Alongside the index we report its four components separately, together with other observable margins of driver behavior: average and maximum speed, kilometers, hours, revenue, pay, and alert rates per hour. Appendix Table A7 shows that these tracker-based measures line up with human-coded unsafe driving from validation rides. They also track crash risk: at matched points in the operating day, vehicle-days that ended in an accident were speeding about 2.4 times as often as other vehicle-days (Appendix Figure A2), although with few accidents this gap is not statistically significant.

5 Results

Our baseline specification regresses each daily outcome on an indicator for active eligibility for cash incentives, controlling for mileage and including bus and calendar-day fixed effects:

$$y_{it} = \beta \, \mathrm{ActiveCash}_{it} + \gamma \, \mathrm{Mileage}_{it} + \alpha_i + \delta_t + \varepsilon_{it}, \tag{1}$$

where i indexes buses, t indexes calendar days, α_i are bus fixed effects, and δ_t are calendar-day fixed effects. We cluster standard errors by bus.[8]

5.1 Is it possible to incentivize drivers to improve safety?

We first examine the impact of cash incentives on driver safety. Figure 2 presents the treatment effects (see Table A2 for exact point estimates). Panel A shows that the incentives improve a composite safety index by 0.096 standard deviations. This result is driven

[7] For two of our eleven routes, carrying 7 and 3 sample buses, we cannot construct the index as pre-specified: the route's first-two-week pre-period contains no events for at least one alert type (one route is essentially eventless, the other never registers over-acceleration), leaving the normalizing baseline undefined. We drop these two routes (10 buses) from the analysis throughout. Appendix Table A3 reproduces Table A2 for all eleven routes, assigning a zero contribution where a route has no baseline events for an alert type, and the estimates are very similar.

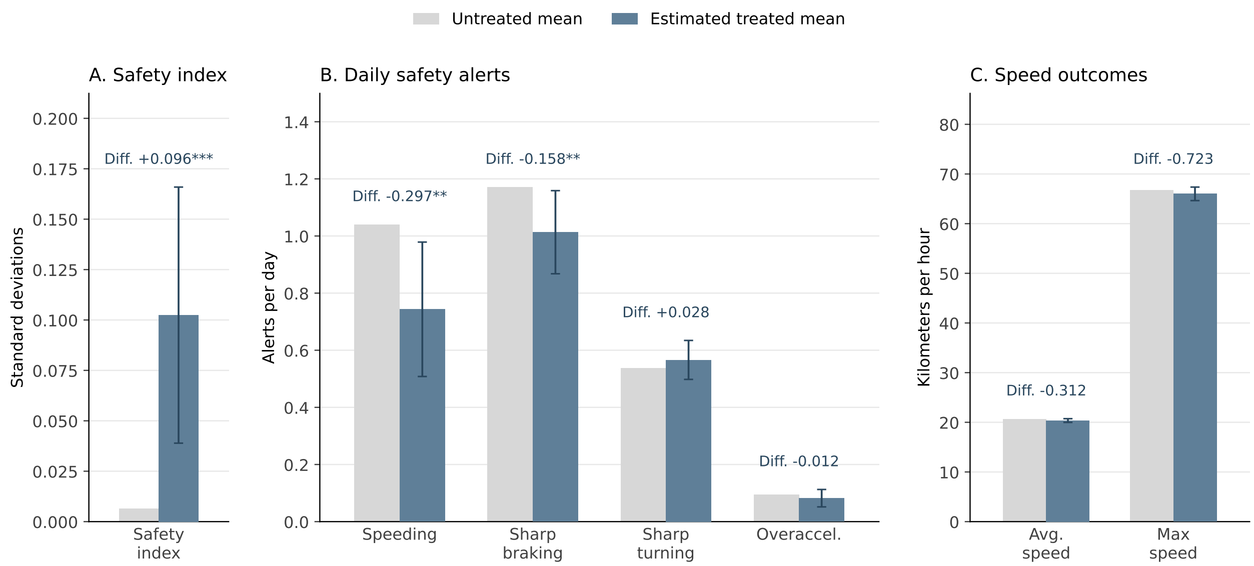
[8] This specification follows the pre-analysis plan, with one exception: we use bus fixed effects instead of route fixed effects. We originally pre-specified route fixed effects because (Kelley et al., 2024) evaluated an information intervention for which no pre-treatment outcome data were available, making bus fixed effects

by two margins, visible in Panel B: speeding falls by 0.297 alerts per day (about 29%) and sharp braking by 0.158 alerts per day (about 13%), both relative to a control mean of roughly one alert per day. The coefficients on overacceleration and sharp turning are small and imprecise. The absence of clear effects on these behaviors is consistent with overacceleration and sharp turning behaviors being noisier, rarer, or less tightly linked to the payout rule (they received less weight in the incentive calculation). Finally, we find little change in average or maximum speed (Panel C). This is perhaps unsurprisingly, as these are relatively coarse measures of driving behavior: average speed aggregates behavior over an entire day and may therefore be insensitive to small behavioral adjustments, while maximum speed captures only a single observation and may similarly fail to detect more subtle changes in driving patterns.

The central result is that raising the cost of unsafe driving (or, analogously, rewarding safer driving) does improve safety even in an environment where enforcement is typically weak. In the language of the model, the direct effect of a higher enforcement price, $\pi_t \tau$, is first-order even though enforcement is temporary and incomplete. The margins that move are the ones drivers can plausibly adjust in real time when unsafe driving becomes more expensive: avoiding high-speed episodes and braking less abruptly.

infeasible. In this study, by contrast, we observe pre-treatment outcomes, allowing us to incorporate bus fixed effects and substantially increase statistical power. The estimated effects are qualitatively similar when we instead use route fixed effects, but, unsurprisingly, the estimates are noisier and largely statistically insignificant. We also report our main results using both bus fixed effects and route-by-day fixed effects in Appendix Table A4, and we show that results are similar.

Figure 2: Main treatment effects in levels



Notes: This figure presents the impact of our treatment on driver's safety behavior. Panel A focuses on the overall composite safety index, for which higher values indicate safer driving. Panel B focuses on the four daily safety alerts that make up the index (speeding, sharp braking, sharp turning, and over-acceleration), and Panel C focuses on speed outcomes (average and maximum speed). The gray bars report untreated means in the Appendix Table A2 estimation sample. The blue bars add the regression coefficient from Appendix Table A2 to the untreated mean, so they can be read as estimated treated means under the regression specification. Whiskers show 95 percent confidence intervals around that estimated treated mean. Text above each outcome reports the treatment-control difference from Appendix Table A2; stars denote statistical significance of that difference.

Next, we examine whether drivers bear measurable costs from driving more safely by estimating equation (1) for each operating margin we have in our data. Table 1, Panel A reports business and labor-supply outcomes: drivers work about 43 minutes more (6.8%), drive 3 kilometers further (2.5%), and earn 118 KES less (1.7%), though only the change in hours is statistically significant. Driver pay and repair costs remain unchanged. Drivers thus respond primarily by working moderately longer hours, while revenue and pay are essentially intact.

Table 1: Observable displacement margins

Panel A: Labor supply and business outcomes					
	Kilometers	Hours worked	Revenue	Driver pay	Repair cost
Active cash incentive	3.1	0.72**	-118	9	-42
	(3.0)	(0.29)	(137)	(26)	(41)
Control mean	124.0	10.6	7,117	999	473
Observations	37,353	37,229	19,378	19,373	15,375
Panel B: Temporal and spatial displacement (% of daily driving)					
	Night	Evening	Off-route	Off-road	Out of metro
Active cash incentive	0.62	-0.32	0.12	-0.03	-0.28
	(0.78)	(0.37)	(0.83)	(0.27)	(0.26)
Control mean	7.05	12.39	14.33	2.56	1.37
Observations	36,735	36,735	36,128	36,262	36,262

Notes: Panel A of this table demonstrates how treatment affects the daily operating margins we collect in the daily panel for drivers. Panel B reports whether any displacement occurs from the 30-second panel: the share of active driving between 21:00 and 05:00 local time (night) and between 19:00 and 21:00 (evening), and the share of moving observations more than 1 km from the bus's matched route shape (off-route), with vertical-acceleration shocks above 0.25g (off-road), and outside the Nairobi metropolitan area (out of metro). Each column reports a separate OLS regression on the active-treatment sample, excluding post-treatment bus-days. All specifications include bus and calendar-day fixed effects with standard errors clustered by bus.

5.2 What explains better safety at no significant cost?

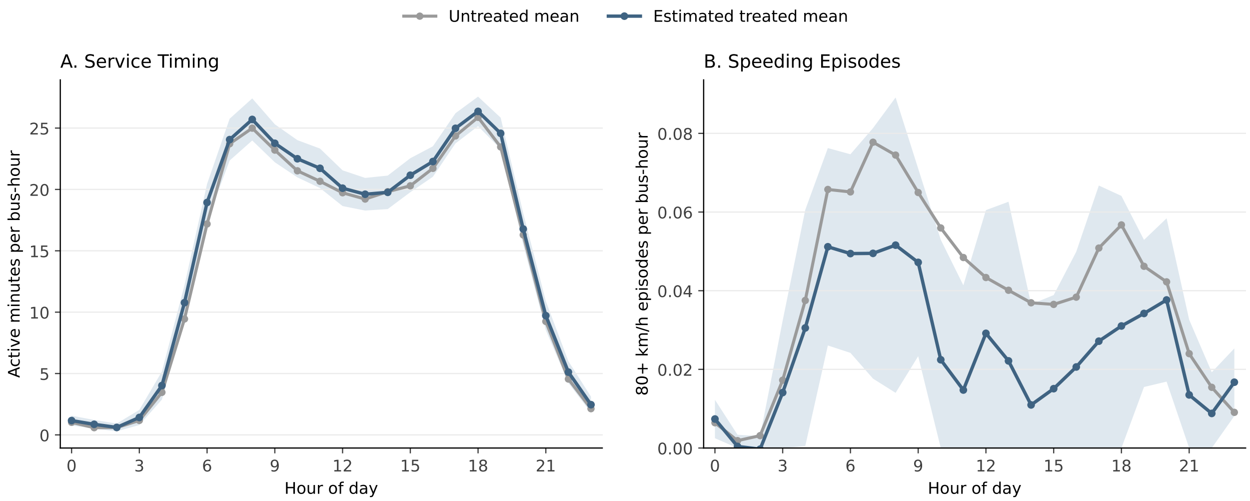
Two scenarios are consistent with these headline results. In the first, drivers' earnings are unchanged because they can drive safely without sacrificing the number of passengers they pick up or the revenue they earn. In the second, earnings are unchanged because drivers comply during the monitored daytime hours but compensate elsewhere—driving less safely, and possibly more, once the payout has been determined. The distinction matters: our safety measures are daily averages, so even if drivers appear safer on the whole, sufficiently dangerous nighttime driving would undo much of the welfare gain we are trying to capture.

We first check for displacement across time and space using the daily panel. Off-route and out-of-metro driving capture the extent to which drivers leave their licensed routes for less-maintained roads to skip traffic, behaviors that are associated with substantially higher costs for owners (Kelley et al., 2024). Panel B of Table 1 shows that the shares of driving at night, in the evening, off-route, off-road, and outside the metro are all unchanged, so drivers do not shift unsafe driving to less-monitored times or places at the daily level. To investigate within-day displacement more finely, we leverage the

30-second tracker data and estimate an hour-specific variant of equation (1), interacting our indicator for active cash treatment with clock-hour indicators. Figure 3 shows two patterns. Panel A plots a clock-hour activity profile: active minutes for each hour of the day, with inactive or unobserved bus-hours coded as zero. We find no evidence that cash incentives shift when drivers work or take buses off the road. The treated and control profiles are nearly identical throughout the day. Panel B instead shows that treated buses trigger fewer speeding episodes (runs of consecutive 30-second records at or above 80 km/h) throughout the operating day, with no systematic nighttime increase. Around 7–8 a.m., the treated profile is roughly 30 percent below the control profile, and the proportional gap is largest in the late morning and mid-afternoon, when episodes fall by half or more.

Taken together, these results suggest that incentive schemes can be an effective way to improve safety in this environment. Drivers respond by driving more safely without any measurable cost to their earnings. This is encouraging from a policy standpoint, since a scheme of this kind is relatively straightforward to implement in practice. It uses off-the-shelf telematics, calibrates against each driver's own history, and requires no new monitoring infrastructure beyond what insurers and fleet operators already deploy.

Figure 3: Within-day operation and speeding



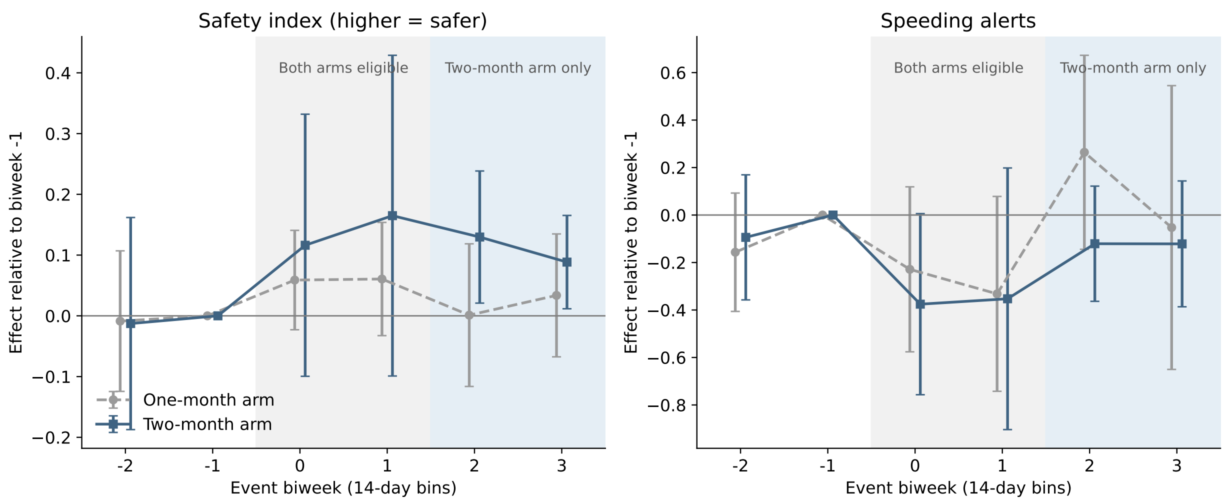
Notes: Hour-specific treatment effects from the 30-second tracker backups for the 212 analysis-sample buses that can be linked one-to-one to the raw telemetry using day-level speed profiles. Panel A reports untreated means and estimated treated means for active minutes per bus-hour, where missing tracker observations in a bus-hour are coded as zero activity; it can therefore be read as a same-time-of-day activity profile. Panel B reports untreated means and estimated treated means for speeding episodes per bus-hour, where an episode is a run of consecutive 30-second records at or above 80 km/h and unobserved bus-hours are coded as zero. In both panels, the gray line reports untreated means and the blue line adds the hour-specific regression coefficient to the untreated mean; the ribbon shows the 95 percent confidence interval around that estimated treated mean. All regressions include bus and calendar-day fixed effects with standard errors clustered by bus.

5.3 Can exposure create persistence?

If cash incentives can promote safe driving without imposing measurable costs on drivers' earnings, an important question is whether drivers eventually internalize these habits and continue driving safely after the incentives end. We conduct two analyses to test for persistence. First, we compare outcomes for buses assigned to the one-month cash incentive arm once their incentives end, with those assigned to the two-month cash incentive arm, which continues to receive payments for an additional month. If outcomes in the one-month arm begin to diverge from those in the two-month arm, we interpret this as evidence that the effects do not persist. Conversely, if outcomes remain similar after payments stop, we interpret this as evidence of persistence. Second, we run an event-study specification by aligning each bus to its own treatment-end date to test if longer exposure leads to greater persistence.

Figure 4 plots treatment effects over time for the one-month and two-month treatment arms for the overall safety index (left panel) and speeding alerts (right panel). As documented above, both treatment arms show improvements in safety and reductions in speeding once incentives are introduced, although precision declines somewhat because we are splitting the sample and this reduces statistical power. More importantly, to test for persistence, we focus on the period after incentives end for the one-month treatment arm (shown in the shaded blue region). We find that buses in the one-month arm revert toward baseline behavior once payments stop. This pattern suggests that safer driving behaviors do not persist after incentives are removed.

Figure 4: Dynamic effects around treatment start



Notes: This figure presents the event-study coefficients from regressions of the safety index and daily speeding on biweekly (14-day) event-time indicators for the one-month and two-month treatment arms, omitting the two-week period immediately before treatment start. All regressions control for kilometers driven and include bus and calendar-day fixed effects with standard errors clustered by bus. The gray region marks periods when both treatment arms are eligible for bonuses; the blue region marks periods when only the two-month arm remains eligible. Unlike Figure 3, which uses matched 30-second tracker backups, this figure uses the daily panel to show treatment dynamics and fade-out.

To test if longer exposure leads to greater persistence, Appendix Table A8 aligns each bus to its own treatment-end date, the day its incentive eligibility expired, and asks how driving evolves in the weeks that follow. Restricting to treated buses, we estimate an event-time analog of equation (1) that interacts indicators for two-week windows after a bus's incentives end with the two-month arm:

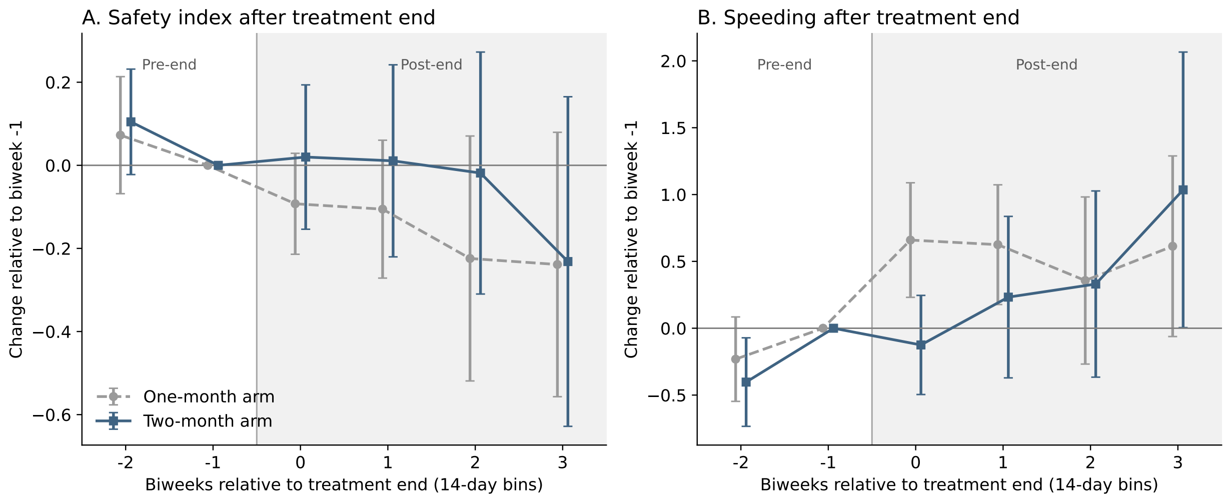
$$y_{it} = \sum_{w} (\beta_w \, \mathrm{Post}^w_{it} + \theta_w \, \mathrm{Post}^w_{it} \times \mathrm{TwoMonth}_i) + \gamma \, \mathrm{Mileage}_{it} + \alpha_i + \delta_t + \varepsilon_{it}, \tag{2}$$

where Post^w_{it} indicates that bus i is in the w-th two-week window after its incentives end, for $w = 1, 2, 3, 4$, and $\mathrm{TwoMonth}_i$ indicates the two-month arm. The omitted reference is the week immediately before each bus's incentives ended, so every coefficient is measured relative to that pre-end week. The base terms β_w capture how the one-month arm changes after its incentives end, while the interactions θ_w capture the two-month arm's *additional* change relative to the one-month arm, so θ_w identifies whether longer exposure builds more durable habits. As in equation (1), α_i and δ_t are bus and calendar-day fixed effects and standard errors are clustered by bus.

Once its incentives end, the one-month arm's alert counts rise immediately, though the composite index reacts more slowly. In the first two weeks after expiration, speeding rises by 0.504 alerts per day and sharp braking by 0.405, both significant, while the safety index falls by only 0.059 standard deviations, not distinguishable from its pre-end level. The index then drifts steadily lower, reaching -0.326 standard deviations by weeks 5–6 and -0.390 by weeks 7–8, though both estimates remain imprecise.

The interaction terms compare the two arms over the same window, so a negative speeding coefficient means the two-month arm deteriorated less than the one-month arm. Longer exposure clearly cushions the reversal in speeding: in the first two weeks after incentives end, the two-month arm speeds 0.790 fewer times per day than the one-month arm ($t = 2.8$), and its net change of $0.504 - 0.790 = -0.286$ alerts per day leaves it slightly safer than before its bonus ended. Similarly, in the first two weeks after the incentives end, the two-month arm brakes less sharply 0.079 fewer times per day, though not statistically significant. On the composite safety index the two-month arm scores 0.198 standard deviations higher than the one-month arm over the same weeks, which points in the same direction as the speeding results but not statistically significant ($t = 1.4$). So the evidence that longer exposure temporarily helps drivers hold onto safer driving is clear for speeding and only suggestive for the index as a whole. Either way, the advantage is short-lived. Figure 5 traces out a wedge between the treatment group: it is substantial and stable throughout the first few weeks but the two arms converge by weeks 7–8.

Figure 5: Persistence after treatment end by randomized duration



Notes: Event-study coefficients from treated-bus regressions on biweekly (14-day) event-time indicators, aligning each bus by its treatment end date and omitting the two-week period immediately before incentives expired. The two panels show the safety index and speeding. The one-month and two-month arms were randomly assigned among treated buses, so differences after treatment end isolate whether longer exposure created additional persistence. All regressions control for kilometers driven, include bus and calendar-day fixed effects, and cluster standard errors by bus.

Taken together, the post-treatment evidence points to modest and short-lived habit formation. In the language of the model, the pattern is consistent with complementarity that is positive but too weak to be self-sustaining: longer exposure shifts the habitual level further, but once pricing ends behavior reverts toward the old habit as the stock decays.

5.4 Illustrative Costs and Benefits

In this section we consider whether the safety gains are large enough to justify the intervention's transfer costs. First, we consider the cost. We estimate total bonus transfers in the analysis sample at approximately 1,317,000 KES, about two thirds of the maximally possible total transfer (2,062,000 KES). The point estimate on repair costs implies the intervention saved owners roughly 12% of this amount in vehicle maintenance, bringing the net transfer cost down to about 1,161,000 KES. Next, we consider how many statistical deaths the intervention would need to avert to break even. Standard developing-country estimates value a statistical life at 30 to 70 million KES (Milligan et al., 2014). Setting the net cost against this range, the intervention would need to avert between 0.02 $(= 1,161,000/70,000,000)$ to 0.04 $(= 1,161,000/30,000,000)$ statistical deaths over the study window to cover its cost.

Fatalities are rare, whereas accidents are more common. To convert fatalities into accidents note that in our sample 1 out of every 28 accidents was fatal, so each accident averted is worth, on average, 1/28 of a fatality averted. Therefore, the intervention would need to avert roughly 0.5 to 1.1 accidents over the study window to break even. We can now compare this break-even range to the estimated number of accidents averted by the cash treatment. The accident regression in Appendix Table A6 yields a point estimate of 0.00052 fewer accidents (not statistically significant) per bus-day during active treatment. Scaled across the 3,718 active treated bus-days in the sample, the intervention is estimated to have averted about 1.9 accidents in total ($0.00052 \times 3,718 \approx 1.9$). This point estimate is imprecise as it comes from a regression on a sparse outcome, and the true number could be meaningfully higher or lower. Still, the central estimate sits well above the 0.5 to 1.1 break-even range, suggesting the intervention is plausibly cost-effective on safety grounds alone, even before counting nonfatal injuries and property damage, which would only strengthen the case.

6 Conclusion

Pricing unsafe driving can improve safety in informal transit, but its value depends on whether drivers simply move risk into less-monitored periods and whether safer behavior survives once the incentive lapses. In this experiment, active cash incentives reduced speeding and sharp braking, improved a composite safety index, and did so without obvious reductions in daily operating activity or clear within-day displacement into other hours. The effects appear quickly, and a longer two-month exposure leaves some short-run persistence after incentives end, but the gains still fade once the contract reverts.

The broader implication is that pricing unsafe driving is effective but not self-sustaining. When driver compensation prices unsafe behavior directly, passengers receive safer rides even without strong evidence of offsetting displacement across the day. That logic extends to the growing set of safety policies that price or constrain risky driving imperfectly, including point-to-point cameras, telematics-based insurance or fleet contracts, and in-vehicle speed assistance. The ubiquity of tracking devices makes temporary cash incentives a plausible policy for especially dangerous or weakly enforced routes. But if the underlying contract returns to business as usual once bonuses expire, those gains may only last if the policy is sustained for much longer than a couple of months.

Viewed through the lens of behavioral labor supply, the same mechanisms therefore cut in opposite directions for policy: substitution into low-enforcement states is not strong enough to undo the active treatment effect, but habit formation is also not strong enough to sustain safety once pricing disappears, at least not after two months. For low-enforcement settings, that means sustained pricing of unsafe driving still promises substantial benefits, because durable improvement likely requires contracts or enforcement that keep risky driving priced on a continuing basis.

References

Ashenfelter, Orley, Kirk Doran, and Bruce Schaller (2010): "A Shred of Credible Evidence on the Long-Run Elasticity of Labour Supply," *Economica*, Vol. 77, pp. 637–650.

Becker, Gary S. and Kevin M. Murphy (1988): "A Theory of Rational Addiction," *Journal of Political Economy*, Vol. 96, pp. 675–700.

Behrens, Roger, Dorothy McCormick, and David Mfinanga (2015): *Paratransit in African Cities: Operations, Regulation and Reform*: Routledge.

Brown, Gabriel, Morgan Hardy, Isaac Mbiti, Jamie McCasland, and Isabelle Salcher (2024): "Can Financial Incentives to Firms Improve Apprenticeship Training? Experimental Evidence from Ghana," *American Economic Review: Insights*, Vol. 6, pp. 120–36.

Camerer, Colin, Linda Babcock, George Loewenstein, and Richard Thaler (1997): "Labor Supply of New York City Cabdrivers: One Day at a Time," *The Quarterly Journal of Economics*, Vol. 112, pp. 407–441.

Cefala, Luisa, Supreet Kaur, Heather Schofield, and Yogita Shamdasani (2024): "Habit Formation in Labor Supply," Working Paper.

Cervero, Robert and Aaron Golub (2007): "Informal Transport: A Global Perspective," *Transport Policy*, Vol. 14, pp. 445–457.

Charness, Gary and Uri Gneezy (2009): "Incentives to Exercise," *Econometrica*, Vol. 77, pp. 909–931.

Chou, Yuan K. (2002): "Testing Alternative Models of Labour Supply: Evidence from Taxi Drivers in Singapore," *The Singapore Economic Review*, Vol. 47, pp. 17–47.

Crawford, Vincent P. and Juanjuan Meng (2011): "New York City Cab Drivers' Labor Supply Revisited: Reference-Dependent Preferences with Rational-Expectations Targets for Hours and Income," *American Economic Review*, Vol. 101, pp. 1912–1932.

De Pauw, Ellen, Stijn Daniels, Tom Brijs, Elke Hermans, and Geert Wets (2014): "Behavioural Effects of Fixed Speed Cameras on Motorways: Overall Improved Speed Compliance or Kangaroo Jumps?" *Accident Analysis & Prevention*, Vol. 73, pp. 132–140.

Dupas, Pascaline, Jonathan Robinson, and Santiago Saavedra (2020): "The Daily Grind: Cash Needs and Labor Supply," *Journal of Economic Behavior & Organization*, Vol. 177, pp. 399–414.

Farber, Henry S. (2008): "Reference-Dependent Preferences and Labor Supply: The Case of New York City Taxi Drivers," *American Economic Review*, Vol. 98, pp. 1069–1082.

——— (2015): "Why You Can't Find a Taxi in the Rain and Other Labor Supply Lessons from Cab Drivers," *The Quarterly Journal of Economics*, Vol. 130, pp. 1975–2026.

Habyarimana, James and William Jack (2011): "Heckle and Chide: Results of a Randomized Road Safety Intervention in Kenya," *Journal of Public Economics*, Vol. 95, pp. 1438–1446.

Hauer, Ezra, F. Jorn Ahlin, and James S. Bowser (1982): "Speed Enforcement and Speed Choice," *Accident Analysis & Prevention*, Vol. 14, pp. 267–278.

Høye, Alena (2014): "Speed Cameras, Section Control, and Kangaroo Jumps—A Meta-Analysis," *Accident Analysis & Prevention*, Vol. 73, pp. 200–208.

Hussam, Reshmaan, Atonu Rabbani, Giovanni Reggiani, and Natalia Rigol (2022): "Rational Habit Formation: Experimental Evidence from Handwashing in India," *American Economic Journal: Applied Economics*, Vol. 14, pp. 1–41.

Inada, Hitoshi, Tomohisa Nagata, Odgerel Chimed-Ochir, Takako Yamagami, Takuji Aoyama, and Tetsuya Mizoue (2022): "Effect of Annual Road Safety Publicity and Enforcement Campaign on Road Fatalities in Japan," *Journal of Epidemiology and Community Health*, Vol. 76, pp. 146–151.

Johnson, Ryan M., David H. Reiley, and Juan Carlos Muñoz (2015): "The War for the Fare: How Driver Compensation Affects Bus System Performance," *Economic Inquiry*, Vol. 53, pp. 1401–1419.

Kelley, Erin M., Gregory Lane, and David Schönholzer (2024): "Monitoring in Small Firms: Experimental Evidence from Kenyan Public Transit," *American Economic Review*, Vol. 114, pp. 3119–3160.

Lazear, Edward P. (2006): "Speeding, Terrorism, and Teaching to the Test," *The Quarterly Journal of Economics*, Vol. 121, pp. 1029–1061.

Macharia, W. M., E. K. Njeru, F. Muli-Musiime, and V. Nantulya (2009): "Access to Quality Care in the Road Safety Context: The Case of Road Traffic Injuries in Kenya," *African Health Sciences*, Vol. 9, pp. 118–124.

Malekpour, Mohammad-Reza, Seyyed-Hadi Ghamari, Erfan Ghasemi, Seyedamirhossein Hejaziyeganeh, Mohsen Abbasi-Kangevari, Kavi Bhalla, Nazila Rezaei, Saeid Shahraz, Arezou Dilmaghani-Marand, Seyed Taghi Heydari, Negar Rezaei, Kamran B. Lankarani, and Farshad Farzadfar (2023): "The Effect of Real-Time Feedback and Incentives on Speeding Behaviors Using Telematics: A Randomized Controlled Trial," *Accident Analysis & Prevention*, Vol. 191, p. 107216.

McCormick, Dorothy, Winnie Mitullah, Preston Chitere, Risper Orero, and Marilyn Ommeh (2013): "Paratransit Business Strategies: A Bird's-Eye View of Matatus in Nairobi," *Journal of Public Transportation*, Vol. 16.

Michuki, John (2003): "The Traffic Act."

Milligan, Craig, Andreas Kopp, Said Dahdah, and Jeannette Montufar (2014): "Value of a Statistical Life in Road Safety: A Benefit-Transfer Function with Risk-Analysis Guidance Based on Developing Country Data," *Accident Analysis & Prevention*, Vol. 71, pp. 236–247.

Mutongi, Kenda (2006): "Thugs or Entrepreneurs? Perceptions of Matatu Operators in Nairobi, 1970 to the Present," *Africa: Journal of the International African Institute*, Vol. 76, pp. 549–568.

——— (2017): *Matatu: A History of Popular Transportation in Nairobi*: University of Chicago Press.

Mwende, Sharon (2025): "New City Hall rules to limit number of matatus per route," The Star, May, Accessed June 5, 2026.

Rasmussen, Jacob (2012): "Inside the System, Outside the Law: Operating the Matatu Sector in Nairobi," *Urban Forum*, Vol. 23, pp. 415–432.

Raynor, Nicolas J. and Tolib Mirzoev (2014): "Understanding Road Safety in Kenya: Views of Matatu Drivers," *International Health*, Vol. 6, pp. 242–248.

Reback, Randall (2008): "Teaching to the Rating: School Accountability and the Distribution of Student Achievement," *Journal of Public Economics*, Vol. 92, pp. 1394–1415.

Reed, Rachel C. (2018): "Transportation Turned Performance Art: Nairobi's Matatu Crews," *New York Times*.

Royer, Heather, Mark Stehr, and Justin Sydnor (2015): "Incentives, Commitments, and Habit Formation in Exercise: Evidence from a Field Experiment with Workers at a

Fortune-500 Company," *American Economic Journal: Applied Economics*, Vol. 7, pp. 51–84.

Stevenson, Mark, Anthony Harris, Jasper S. Wijnands, and Duncan Mortimer (2021): "The Effect of Telematic Based Feedback and Financial Incentives on Driving Behaviour: A Randomised Trial," *Accident Analysis & Prevention*, Vol. 159, p. 106278.

Vaa, Truls (1997): "Increased Police Enforcement: Effects on Speed," *Accident Analysis & Prevention*, Vol. 29, pp. 373–385.

WHO, Geneva (2015): "Global Status Report on Road Safety,"Technical report, World Health Organization, Geneva.

WHO, Geneva (2020): "Road Traffic Injuries: Key Facts,"Technical report, World Health Organization.

Williams, Sarah, Adam White, Peter Waiganjo, Daniel Orwa, and Jacqueline Klopp (2015): "The Digital Matatu Project: Using Cell Phones to Create an Open Source Data for Nairobi's Semi-Formal Bus System," *Journal of Transport Geography*, Vol. 49, pp. 39–51.

Zou, Eric Yongchen (2021): "Unwatched Pollution: The Effect of Intermittent Monitoring on Air Quality," *American Economic Review*, Vol. 111, pp. 2101–2126.

Appendix

Appendix Table A1: Balance on cash-incentive assignment

	Control	One-month arm			Two-month arm		
	Mean	Mean	Diff. (s.e.)	p-value	Mean	Diff. (s.e.)	p-value
Driver and experiment characteristics							
Owner monitoring treatment	0.500	0.585	0.085 (0.091)	0.347	0.528	0.028 (0.097)	0.774
Driver age	36.07	36.71	0.64 (1.38)	0.645	35.17	-0.90 (1.29)	0.482
Driver education	2.48	2.37	-0.12 (0.10)	0.257	2.36	-0.12 (0.14)	0.387
Driver experience	8.10	8.85	0.76 (1.29)	0.556	7.22	-0.87 (0.91)	0.339
Pre-treatment operations							
Kilometers	119.2	118.9	-0.3 (8.3)	0.975	118.5	-0.7 (10.5)	0.949
Hours worked	10.35	10.41	0.05 (0.52)	0.920	10.84	0.48 (0.75)	0.517
Revenue target	2955	2820	-135 (128)	0.290	2919	-36 (117)	0.757
Revenue	6919	7423	505 (392)	0.197	6766	-152 (442)	0.730
Driver pay	980	996	16 (70)	0.815	989	9 (67)	0.892
Pre-treatment safety							
Avg. speed	20.491	20.186	-0.305 (1.004)	0.762	20.266	-0.224 (1.115)	0.841
Speeding	0.941	0.746	-0.195 (0.265)	0.462	1.314	0.372 (0.388)	0.337
Sharp braking	1.085	0.989	-0.095 (0.262)	0.717	1.352	0.267 (0.344)	0.438
N buses	126	41			36		

Notes: Each column reports bus-level means in the later cash-treatment analysis sample. "Owner monitoring treatment" records whether the bus's owner was assigned access to driver monitoring reports in Kelley et al. (2024). For treated buses, pre-treatment rows average over up to 30 observed positive-mileage bus-days immediately preceding the later cash-treatment start date. For control buses, we assign placebo start dates using the empirical distribution of treated start dates within route whenever possible, and otherwise from the full treated distribution, and apply the same 30-observed-day rule. Differences, standard errors, and p-values come from separate cross-sectional regressions of each row variable on the arm indicator, comparing each treatment arm to the control group, with heteroskedasticity-robust (HC3) standard errors.

Appendix Table A2: Driver cash incentives and daily safety

		Index components				*Other safety outcomes*			
	Safety index	Speeding	Sharp braking	Sharp turning	Overaccel.	Avg. speed	Max speed	Speeding per hour	Braking per hour
Active cash incentive	0.096***	-0.297**	-0.158**	0.028	-0.012	-0.312	-0.723	-0.026**	-0.017***
	(0.032)	(0.120)	(0.074)	(0.035)	(0.016)	(0.192)	(0.694)	(0.011)	(0.006)
Control mean	0.007	1.041	1.171	0.538	0.094	20.665	66.728	0.102	0.108
Observations	37,353	37,353	37,353	37,353	37,353	37,353	37,353	35,867	35,867

Notes: This table presents the impact of our treatment on drivers' safety behavior. Column 1 reports the pre-specified composite safety index: a weighted average of route-normalized daily event rates for the four alert types (weights of 1/3 on speeding and sharp braking, 1/6 on sharp turning and over-acceleration), negated and standardized using the control-group mean and standard deviation so that higher values indicate safer driving. Columns 2–5 report the four alert components of the index: speeding, sharp braking, sharp turning, and over-acceleration. Columns 6–9 report other safety outcomes: average speed, maximum speed, speeding per hour, and braking per hour. The speeding- and braking-per-hour columns normalize alerts by hours worked and are estimated on bus-days with at least one recorded work hour. Each column reports a separate OLS regression on the active-treatment sample, excluding post-treatment bus-days. "Active cash incentive" equals one when a treated driver is currently eligible for payments. All specifications control for kilometers driven and include bus and calendar-day fixed effects. Standard errors (in parentheses) are clustered at the bus level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Appendix Table A3: Driver cash incentives and daily safety: all eleven routes

		Index components				*Other safety outcomes*			
	Safety index	Speeding	Sharp braking	Sharp turning	Overaccel.	Avg. speed	Max speed	Speeding per hour	Braking per hour
Active cash incentive	0.085***	-0.266**	-0.155**	0.031	-0.010	-0.239	-0.633	-0.023**	-0.017***
	(0.031)	(0.111)	(0.070)	(0.033)	(0.014)	(0.182)	(0.649)	(0.010)	(0.006)
Control mean	0.007	1.012	1.144	0.520	0.091	20.497	66.517	0.099	0.105
Observations	39,072	39,072	39,072	39,072	39,072	39,072	39,072	37,530	37,530

Notes: This table reproduces Table A2 on the full sample of all eleven routes, including the two routes (10 buses) excluded from the main analysis because the pre-specified route baseline cannot be computed for at least one alert type. For those routes, an alert type with no events in the first-two-week pre-period is assigned a zero contribution to the index. The specification is otherwise identical to Table A2: each column is a separate OLS regression on the active-treatment sample excluding post-treatment bus-days, controlling for kilometers driven, with bus and calendar-day fixed effects and standard errors clustered by bus. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Appendix Table A4: Robustness to route-by-day fixed effects

		Index components				*Other safety outcomes*			
	Safety index	Speeding	Sharp braking	Sharp turning	Overaccel.	Avg. speed	Max speed	Speeding per hour	Braking per hour
Active cash incentive	0.073*	-0.285**	-0.122	-0.026	-0.011	-0.368*	-1.041	-0.025**	-0.014**
	(0.038)	(0.133)	(0.074)	(0.036)	(0.014)	(0.195)	(0.733)	(0.012)	(0.006)
Control mean	0.007	1.041	1.171	0.538	0.094	20.665	66.728	0.102	0.108
Observations	37,353	37,353	37,353	37,353	37,353	37,353	37,353	35,867	35,867

Notes: Each column reports a separate OLS regression on the active-treatment sample, excluding post-treatment bus-days. Relative to Appendix Table A2, the specification replaces the calendar-day fixed effects with route-by-day fixed effects, so treatment effects are identified by comparing treated and untreated buses operating on the same route on the same day. All specifications control for kilometers driven, include bus fixed effects, and cluster standard errors by bus. The speeding- and braking-per-hour columns normalize alerts by hours worked and are estimated on bus-days with at least one recorded work hour.

Appendix Table A5: Robustness to using only the second cash stint

		Index components				*Other safety outcomes*			
	Safety index	Speeding	Sharp braking	Sharp turning	Overaccel.	Avg. speed	Max speed	Speeding per hour	Braking per hour
Active cash incentive	0.122**	-0.402***	-0.235**	0.049	-0.035**	-0.389	-0.814	-0.032**	-0.022**
	(0.050)	(0.155)	(0.098)	(0.045)	(0.018)	(0.274)	(0.953)	(0.014)	(0.009)
Control mean	0.007	1.041	1.171	0.538	0.094	20.665	66.728	0.102	0.108
Observations	35,774	35,774	35,774	35,774	35,774	35,774	35,774	34,340	34,340

Notes: Each column reports a separate OLS regression, excluding post-treatment bus-days. Relative to Appendix Table A2, the sample is restricted to the second cash stint: for treated buses that received the incentive in two stints, first-stint bus-days are dropped and the active-treatment indicator is defined from the later stint only. All specifications control for kilometers driven, include bus and calendar-day fixed effects, and cluster standard errors by bus.

Appendix Table A6: Recorded accident outcomes

	Any accident	Matatu injuries	Third-party injuries	Fatalities
Active cash incentive	-0.052	-0.028	-0.005	0.008
	(0.080)	(0.025)	(0.005)	(0.008)
Control mean	0.077	0.018	0.003	0.003
Observations	37,353	37,353	37,353	37,353

Notes: Each column reports a separate OLS regression on the active-treatment sample, excluding post-treatment bus-days. Outcomes are merged from the study accident file and include an indicator for any recorded accident, passenger injuries in the matatu, third-party injuries, and fatalities. All outcome variables are multiplied by 100, so coefficients and control means are in percentage points of a bus-day. All specifications control for kilometers driven, include bus and calendar-day fixed effects, and cluster standard errors by bus. The outcome data are sparse, and the fatality column is based on a single recorded fatality in a control bus-day, so these estimates should be interpreted as descriptive scale rather than precise treatment effects.

Appendix Table A7: External validation of tracker safety measures

	Human hard braking	Human speeding	Any human unsafe event
Tracker variable	Brake alerts	Speeding alerts	Safety index
Tracker measure (std.)	0.091**	0.079**	-0.069***
	(0.043)	(0.035)	(0.019)
Outcome mean	0.535	0.183	0.421
Observations	617	617	617
Sample	Human validation bus-days	Human validation bus-days	Human validation bus-days

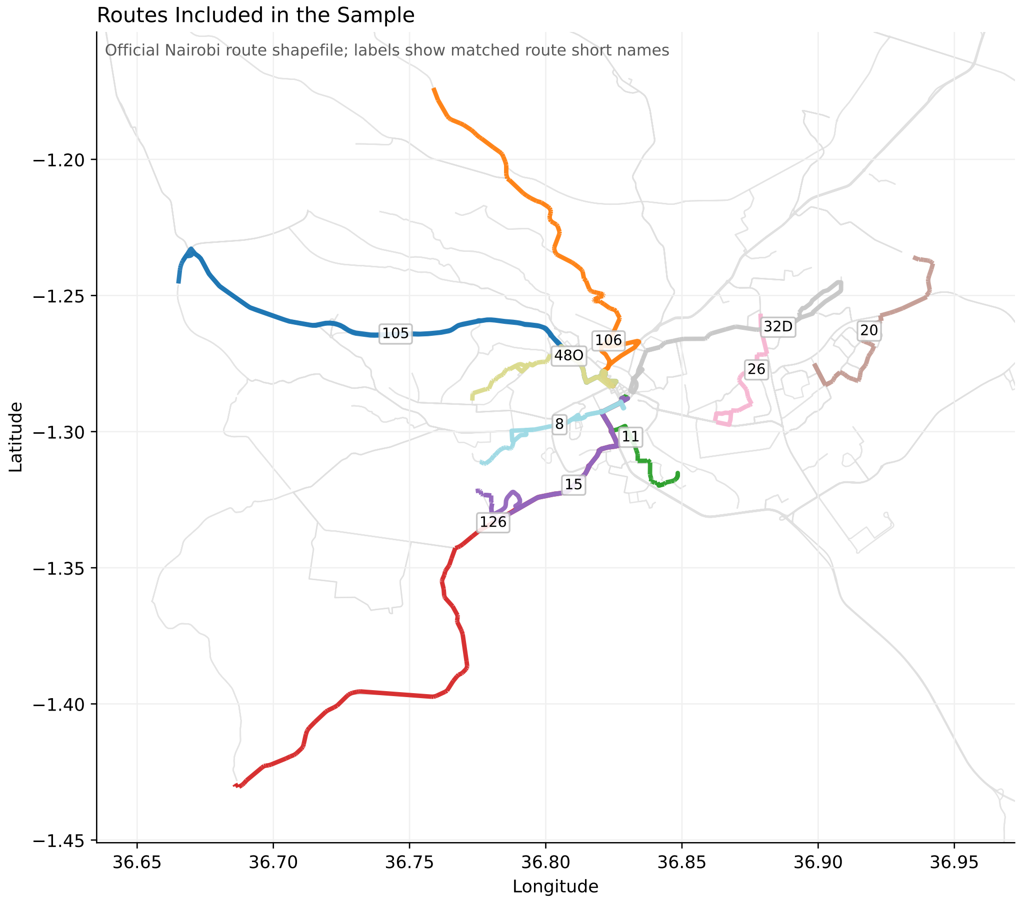
Notes: Each column reports a separate OLS regression with standard errors clustered by bus. The three columns use bus-days from the human-validation rides that can be matched to the daily tracker data by plate and date. The outcomes are the count of human-coded hard-braking events, the count of human-coded speeding events, and an indicator for any human-coded hard braking, speeding, or off-route driving on that bus-day. The row "Tracker variable" identifies the tracker measure used in each column. All tracker variables are standardized to have unit standard deviation within the estimation sample, and all specifications control for mileage.

Appendix Table A8: Persistence by randomized duration after incentives end

	Weeks after treatment end			
	1–2	3–4	5–6	7–8
Safety index (N = 1,608)				
One-month arm (β_w)	-0.059	-0.142	-0.326*	-0.390*
	(0.074)	(0.119)	(0.189)	(0.218)
Two-month arm, additional (θ_w)	0.198	0.212	0.274	0.017
	(0.140)	(0.186)	(0.176)	(0.170)
Speeding (N = 1,608)				
One-month arm (β_w)	0.504***	0.636**	0.481	0.735
	(0.177)	(0.292)	(0.460)	(0.515)
Two-month arm, additional (θ_w)	-0.790***	-0.599	-0.313	0.266
	(0.282)	(0.380)	(0.408)	(0.419)
Sharp braking (N = 1,608)				
One-month arm (β_w)	0.405**	0.420	0.546	0.664
	(0.167)	(0.292)	(0.426)	(0.509)
Two-month arm, additional (θ_w)	-0.079	0.082	-0.200	-0.146
	(0.197)	(0.308)	(0.284)	(0.301)
Sharp turning (N = 1,608)				
One-month arm (β_w)	0.043	0.113	0.353*	0.527**
	(0.118)	(0.150)	(0.202)	(0.260)
Two-month arm, additional (θ_w)	-0.007	-0.153	-0.057	0.058
	(0.175)	(0.217)	(0.280)	(0.228)
Over-acceleration (N = 1,608)				
One-month arm (β_w)	-0.014	0.026	0.117	0.106
	(0.031)	(0.098)	(0.161)	(0.212)
Two-month arm, additional (θ_w)	0.032	-0.003	0.039	-0.011
	(0.043)	(0.050)	(0.073)	(0.074)

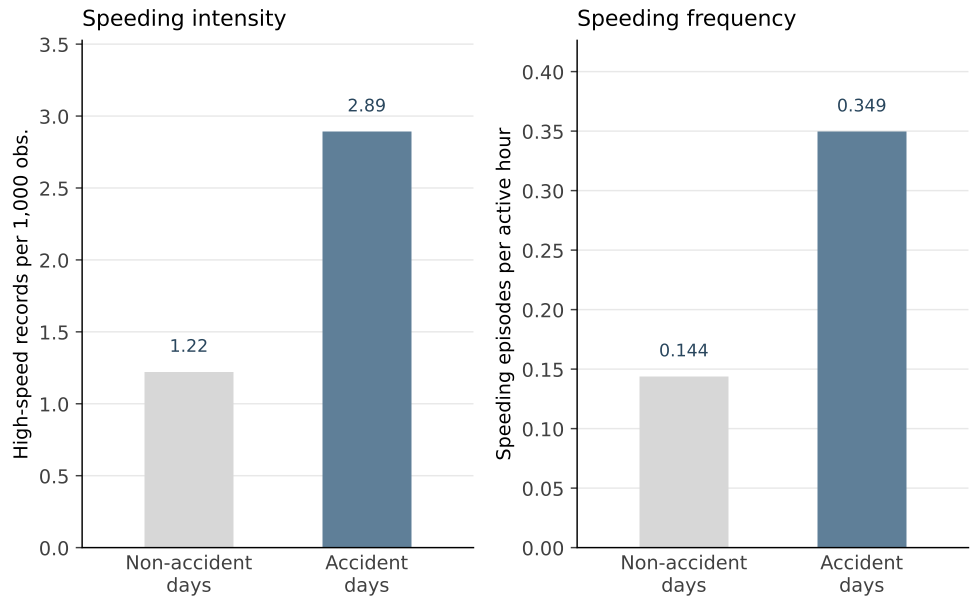
Notes: Treated-bus OLS regressions on bus-days from the week before each arm's treatment end through eight weeks after. Each panel reports the two coefficients of a single regression of one outcome on two-week post-expiration window indicators interacted with the two-month arm, measured relative to the week immediately before incentives expired: "One-month arm (β_w)" is the one-month arm's change after its incentives end, and "Two-month arm, additional (θ_w)" is the two-month arm's additional change relative to the one-month arm (the interaction term in equation (2)). The two-month arm's own net change is $\beta_w + \theta_w$. All specifications control for kilometers driven, include bus and calendar-day fixed effects, and cluster standard errors by bus.

Appendix Figure A1: Sample routes in Nairobi



Notes: Appendix figure mapping the Nairobi routes represented in the driver experiment. The figure plots the official Nairobi matatu route shapefile from the main project. Sample routes are identified by matching each analysis-sample route's linked GPS footprint to the route shapes, using human-validation route numbers when available to discipline the match. Colored lines show the matched sample routes; faint gray lines show the broader route network.

Appendix Figure A2: Speeding and accidents at matched operating hours



Notes: Each vehicle-day is aligned by elapsed operating hour (the first active hour, the second, and so on) using the 30-second tracker data, so that accident days, which end early, are compared with non-accident days over the same operating hours. Bars report the operating-hour-adjusted mean speeding rate by whether the vehicle-day had a recorded accident: high-speed records (80 km/h or above) per 1,000 active records (left, "speeding intensity") and speeding episodes per active hour (right, "speeding frequency"). The comparison rests on 25 accident vehicle-days, so the difference, though large, is not statistically significant.